УДК 004.056

АНАЛИЗ БЕЗОПАСНОСТИ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ОБЪЕКТНО ОРИЕНТИРОВАННОЙ АРХИТЕКТУРЕ УМНОГО ДОМА

А. И. Дубровина, Д. В. Маршаков
Донской государственный технический университет (г. Ростов-на-Дону, Российская Федерация)

Комплексная реализация принципов технической архитектуры в системе домашней автоматизации, обеспечивающей обмен данными между интерфейсом пользователя и устройствами интернета вещей, представляет собой объектно ориентированную архитектуру умного дома. Применение в данной архитектуре глубоких нейронных сетей (ГНС) позволяет упростить и ускорить автоматизированные процессы. Актуальны анализ безопасности и оценка уязвимостей такого рода нейросетевых систем в условиях вредоносных воздействий. В настоящей статье представлен анализ основных видов атак на ГНС и перечислены мероприятия по их предотвращению.

Ключевые слова: умный дом, интернет вещей, искусственные нейронные сети, глубокое обучение, вредоносное глубокое обучение.

SECURITY ANALYSIS OF THE APPLICATION OF NEURAL NETWORKS IN OBJECT-ORIENTED SMART HOME ARCHITECTURE

A.I. Dubrovina, Marshakov D.V.
Don State Technical University (Rostov-on-Don, Russian Federation)

A comprehensive implementation of the principles of technical architecture in a home automation system that provides data exchange between the user interface and IoT-devices is an object-oriented smart home architecture. The use of deep neural networks (DNN) in this architecture makes it possible to simplify and speed up many automated processes. At the same time, it is relevant to analyze the security and assess the vulnerabilities of such neural network systems in terms of the reliability of their practical application with the adversarial deep learning. The paper deals with the analysis of the main types of attacks on the DNN and provides measures to prevent them.

Keywords: smart home, internet of things, artificial neural networks, deep learning, adversarial deep learning.

Введение. В промышленности и домашней автоматизации широко внедряются технологии интернета вещей (англ. internet of things, IoT). Эти решения интегрируют в единую сеть автоматизированное пространство, «умные» вещи, приборы с возможностью удаленного контроля и управления ими без участия человека. Слово «умные» здесь означает «оснащенные датчиками и программным обеспечением для сбора и обмена данными».

Системы такого рода позволяют решать задачи автоматизации технологических процессов масштабных предприятий, сокращения потребления ресурсов, оптимизации времени пользователя, обеспечения комфорта и безопасности на охраняемой территории.

Так, система домашней автоматизации (умный дом) обеспечивает обмен данными между интерфейсом пользователя и IoT-устройствами за счет применения принципов технической архитектуры: наследования, абстракции, полиморфизма, инкапсуляции, отдаления объекта от пользователя. Под объектно ориентированной архитектурой (стилем) умного дома понимают комплексную реализацию этих принципов. На их основе разрабатывается программное

обеспечение по взаимодействию IoT-устройств между собой, обеспечивается получение достоверных сигналов от управляющего контроллера, безошибочное выполнение заданных пользователем сценариев. Основные преимущества данной архитектуры: рекурсивность применения, возможность автоматического тестирования на проникновение (пентестинг), инкапсуляции неоднородных технических решений, простота администрирования. Среди недостатков — специфические программные ошибки при определении класса принадлежности объекта, необходимость полного реструктурирования тела программного кода в случае неудачного подбора класса.

Повышение уровня защищенности охраняемого объекта предполагает обеспечение синхронизации IoT-устройств, а также использование:

— цифровых технологий,

— системы контроля управления доступом (СКУД),

— системы видеоохраны с функцией распознавания биометрических данных и человеческой речи.

Для достижения этих целей необходимо внедрять интеллектуальные системы. Стоит особенно отметить машинное, в том числе глубокое обучение, реализуемое в глубоких нейронных сетях (ГНС). Его применяют, в частности, для обнаружения вторжения в IoT-сеть и выявления вредоносных программ [1], интеллектуального видеонаблюдения за людьми с обеспечением конфиденциальности их личной информации [2]. Актуальны анализ безопасности и оценка уязвимостей таких систем с точки зрения надежности их практического применения в условиях вредоносных воздействий (вредоносного глубокого обучения), представляющих собой целенаправленные воздействия на ГНС для организации ошибок в их поведении [3, 4].

Основная часть

Применение глубоких нейронных сетей. ГНС в системах умного дома играют ключевую роль при решении задач расчетными и аналитическими методами. Данные технологии находят применение в системах «самопентестинг», «мультирум», «безопасность» (СКУД, видеоконтроль), «климат-контроль» и др. Например, в системе «безопасность» владельцы дома могут открывать замки не ключами, а с помощью функции распознавания лица. При этом биометрические данные субъектов доступа предварительно вносятся в базу данных СКУД. Для двумерного либо трехмерного распознавания субъекта задействуют ключевые характеристики его лица, данные сверяются и анализируются.

Для выполнения такой процедуры необходимо предварительное обучение нейронной сети, включающее корректное определение входных данных и репрезентативности обучающей выборки при различных условиях освещенности, положениях головы и пр. В процессе обучения необходимо учесть погрешности и зафиксировать возможные ошибки. Как показывают исследования [5], в результате обучения влияние некоторых факторов обучающей выборки может оказаться незначительным, поскольку весовые коэффициенты связей от таких факторов в процессе обучения стремятся к нулю и могут быть исключены. Затем целесообразно провести повторное обучение, что положительно скажется на целевой ошибке в работе обученных нейронных сетей. Такой подход объясняется большой зашумленностью первоначального набора исходных данных и низкой достоверностью исключенных факторов.

Стоит учесть, что модели глубокого обучения, как правило, имеют дело с миллионами вариантов входных данных, основываясь лишь на небольшом количестве изученных примеров. Это открывает возможность создания уязвимостей ГНС посредством вредоносных воздействий. Такие уязвимости представляют угрозу адекватному функционированию ГНС и определяются как атака на нее.

Виды атак на ГНС. Ниже перечислены потенциально опасные атаки на ГНС [4, 6].

1. «Отравление» исходных данных, основанное на внесении ошибок в обучающую выборку. Это приводит к некорректному выполнению ключевых алгоритмов. Проведение такой атаки достижимо при доступе злоумышленника к обучающей выборке с последующим искажением разметки либо внесением в выборку нетипичных для нее объектов.

2. Создание смысловой (оптической, акустической) иллюзии, подразумевающей замену фактов ложной информацией. Примером такой атаки может служить искажение ранее записанного в СКУД шаблона изображения с целью повышения вероятности возникновения ошибки 1-го рода (т. е. предоставления злоумышленнику доступа на территорию).

Подход к классификации таких атак представлен ниже.

— По желаемому ответу. Система распознавания может принять человека за домашнее животное, и таким образом злоумышленник получит возможность передвигаться по объекту.

— По доступности модели. Атакуя «черный» и «белый» ящики, злоумышленник по реакции системы на действия может заранее определить параметры модели и примеры подлога информации. В такой системе статистические данные будут подменяться в зависимости от целей атакующего.

— По способу подбора помех. Точно рассчитанные модификации исходного изображения позволяют подменить пиксели, тем самым задавая итеративные или однопроходные помехи для генерации вредоносных изображений.

3. Атака идентификации принадлежности. Она основывается на нарушении конфиденциальности данных пользователя и реализуется при сопоставлении статистических показателей, записанных в ГНС, с имеющимися сведениями.

4. Получение исходных данных на основе обученной модели. В этом случае модель-наследник перенимает различные виды конфиденциальной информации, персональные данные, данные счетов, кредитных карт и пр., что может привести к финансовым потерям пользователя.

5. Хищение параметров обучающей модели ГНС. Достаточная осведомленность о структуре ГНС позволяет злоумышленнику создать ее копию с целью выработки и модификации вредоносных входных данных для их последующего применения против целевой системы.

Анализ атак по внесению ошибок в обучение ГНС позволяет сделать вывод о сути проблемы реализации атак. Дело в том, что существующие модели зачастую находятся в незащищенном от удаленного подключения хранилище, облаке или в открытом доступе. Вследствие обработки значительного объема данных при написании программного кода нередко используются старые алгоритмы и статистика, т. к. технически неудобно и финансово невыгодно хранить большие объемы информации разного характера.

Степень устойчивости ГНС к вредоносным воздействиям определяется построением модели угроз целевой системы. При этом нет универсального подхода к оценке модели угроз ГНС. В этом случае ограничиваются использованием:

— эмпирических показателей устойчивости, что справедливо только в отношении конкретной угрозы;

— метрик, полученных теоретическим путем.

Ниже перечислены основные мероприятия по предотвращению атак на ГНС.

— При обучении нейронных сетей запрещается использовать данные непосредственно из физического мира (например, с камер) либо из ненадежных цифровых источников (например, загруженных пользователями на сервер). Такие данные для обучающей выборки может «отравить» третья сторона.

— Важно проверять целостность используемой информации (сопоставлять хеш-коды реального изображения с тем, что используется в настоящий момент).

— Обучение нейросетевых моделей проводится «с нуля», без использования готовых (в том числе предобученных) моделей, т. к. они могут базироваться на ложном алгоритме.

— Обучающая выборка актуализируется с целью последующего повторного обучения ГНС, поскольку доступ даже к некоторым исходным обучающим данным в ряде случаев позволяет создать замещающую модель и получить переносимые вредоносные образы.

— Ограничивается доступ к моделям целевой системы (включающей не только ГНС, но и вычислительную систему в целом, вместе со всеми ее защитными механизмами). В этом случае злоумышленник не изучит ее структуру и свойства, а значит, не воспользуется данными.

— Необходимо обеспечить адекватное соответствие сетевой архитектуры алгоритму выполняемых действий.

— Поведение злоумышленников исследуется с применением модели глубокого обучения на основе статистических данных при различных условиях.

Заключение. Современные алгоритмы машинного обучения позволяют упростить и ускорить множество автоматизированных процессов в объектно ориентированной архитектуре умного дома. Бывает трудно четко определить, по каким признакам данных следует проводить обучение. В этих случаях широко применяется ГНС. С другой стороны, следует отметить повышенную степень уязвимости ГНС к вредоносным входным воздействиям. Различные виды атак на ГНС приводят к ограничению надежности их функционирования, в том числе потенциально неправильному формированию выходных решений. Данное обстоятельство объясняется отчасти тем, что при обучении ГНС сложно или невозможно представить, какие аспекты данных наиболее важны для нейросетевого алгоритма.

Выявление скрытых закономерностей ГНС в обучающей выборке — один из базовых принципов глубокого обучения и одновременно слабое место, поскольку не дает возможности проверить правильность интерпретируемых данных. В связи с этим интересно изучить применение нечеткого моделирования при разработке методов формализации нейросетевых алгоритмов принятия решений, средств извлечения правил из ГНС на этапе их функционирования с целью интерпретации и оценки выходных результатов. Это представляется перспективным направлением исследований.

Библиографический список

1. Lin, T. Deep Learning for IoT / T. Lin // Proc. of the 39th International Performance Computing and Communications Conference. — Austin : IEEE, 2020. — P. 1–4. 10.1109/IPCCC50635.2020.9391558.

2. Chuma, E. L. Internet of Things (IoT) Privacy-Protected, Fall-Detection System for the Elderly Using the Radar Sensors and Deep Learning / E. L. Chuma, L. L. B. Roger, G. G. de Oliveira // Proc. of the International Smart Cities Conference. — Piscataway : IEEE, 2020. — P. 1–4. 10.1109/ISC251055.2020.9239074.

3. Wong, A. Targeted Adversarial Perturbations for Monocular Depth Prediction / A. Wong, S. Cicek, S. Stefano // Cornell University. arXiv : [сайт]. — URL: https://arxiv.org/abs/2006.08602#:~:text=Targeted%20Adversarial%20Perturbations%20for%20Monocular%20Depth%20Prediction,-Alex%20Wong%2C%20Safa&text=We%20study%20the%20effect%20of,perceived%20geometry%20of%20the%20scene. (дата обращения: 02.04.2022).

4. Уорр, К. Надежность нейронных сетей: укрепляем устойчивость ИИ к обману / К. Уорр. — Санкт-Петербург : Питер, 2021. — 272 с.

5. Пирмагомедов, Р. Я. О применении нейросетевых методов для прогнозирования показателей надежности физического канала пассивных оптических сетей / Р. Я. Пирмагомедов // Надежность функционирования и информационная безопасность телекоммуникационных систем железнодорожного транспорта : мат-лы Всерос. науч.-тех. интернет-конф. — Омск : Омский государственный университет путей сообщения, 2013. — С. 209–216.

6. Атаки на нейронные сети: как избежать неприятностей? // smartengines : [сайт]. — URL: https://smartengines.ru/blog/neural-net-attacks/ (дата обращения: 15.03.2022).

Об авторах:

Дубровина Ангелина Игоревна, магистрант кафедры «Вычислительные системы и информационная безопасность» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), adubrovina@yug.gkovd.ru.

Маршаков Даниил Витальевич, к. т. н., доцент кафедры «Вычислительные системы и информационная безопасность» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), кандидат технических наук, доцент, daniil_marshakov@mail.ru.

About the Authors:

Dubrovina, Angelina I., Master's degree student, Department of Computing Systems and Information Security, Don State Technical University (1, Gagarin Square, Rostov-on-Don, 344003, RF), adubrovina@yug.gkovd.ru.

Marshakov, Daniil V., Cand.Sci., Associate professor, Department of Computing Systems and Information Security, Don State Technical University (1, Gagarin Square, Rostov-on-Don, 344003, RF), Cand.Sci., Associate professor, daniil_marshakov@mail.ru