https://orcid.org/0000-0002-9639-3544

УДК 004.457

О.Є. КОВАЛЕНКО*

ІНТЕЛЕКТУАЛІЗАЦІЯ ГРАНИЧНИХ ОБЧИСЛЕНЬ ІНТЕРНЕТУ РЕЧЕЙ

*Національний університет біоресурсів і природокористування України, м. Київ, Україна

Анотація. *Поєднання технологій граничних обчислень та інтернету речей породжує граничний інтернет речей. Граничні обчислення інтернету речей представляють архітектуру граничних обчислень, що інтегруються у систему інтернету речей за допомогою граничного (обчислювального) шлюзу. Граничний обчислювальний шлюз, який об'єднує мережеві, обчислювальні сховища та можливості застосунків, розгортається на межі мережі поблизу пристроїв або джерел даних, щоб він міг надавати сервіси керування пристроями та управління системою інтернету речей у граничних вузлах мережі. Граничний штучний інтелект або «штучний інтелект на межі» означає поєднання граничних обчислень і штучного інтелекту для виконання завдань машинного навчання безпосередньо на взаємопов'язаних граничних пристроях. Одним із напрямів еволюції програмного забезпечення, пов'язаного з граничними обчисленнями, є застосування методів і алгоритмів штучного інтелекту як безпосередньо для обробки інформації, так і для адаптивного програмно керованого розгортання мережевої інфраструктури. Перевагою архітектури граничних обчислень є безпека. Багато компаній не наважуються надсилати конфіденційні дані про свою компанію у хмару та віддають перевагу зберіганню їх у своїх приміщеннях, тим самим зменшуючи ризики кібербезпеки. Конвергенція інформаційних технологій та операційних технологій збільшує потенційну площу атаки і важливо, щоб підприємства зміцнили свої граничні обчислювальні вузли для захисту даних. Швидкий розвиток і поширення технологій штучного інтелекту поширюється у бік граничних обчислень, оскільки штучний інтелект може потенційно забезпечити засоби досягнення властивостей «розумного середовища» гранично-хмарного континууму (edge-cloud continuum). У статті проведено аналіз застосування штучного інтелекту у граничній області інтернету речей. У результаті проведеного аналізу визначено основні переваги та проблеми, пов'язані з граничною областю інтернету речей, та підходи для їх вирішення на основі використання методів штучного інтелекту. Граничні обчислення та граничний штучний інтелект мають різні варіанти використання, але найважливішим є підвищення якості обслуговування для пристроїв інтернету речей.*

Ключові слова: *граничні обчислення, інтернет речей, штучний інтелект, безпека граничних обчислень.*

Abstract. *The combination of edge computing technologies and the Internet of Things creates the edge Internet of Things. Internet of Things edge computing represents the architecture of edge computing integrated into the Internet of Things system with the help of an edge (computing) gateway. This gateway, which integrates network and computing storage and application capabilities, is deployed at the network edge near devices or data sources to provide device management and Internet of Things system management services at network edge nodes. Edge artificial intelligence or «artificial intelligence at the edge» refers to the combination of edge computing and artificial intelligence to perform machine-learning tasks directly on interconnected edge devices. One of the trends in the evolution of software related to edge computing is the use of artificial intelligence methods and algorithms both directly for information processing and for adaptive software-controlled deployment of network infrastructure. The advantage of edge computing architecture is security. Many companies are reluctant to send sensitive company data to the cloud and prefer to keep it on-premises, thereby reducing cybersecurity risks. The convergence of information technology and operational technology increases the potential attack area. Therefore, it is important for enterprises to improve their edge computing nodes to protect their data. The rapid development and spread of artificial intelligence technologies is spreading towards edge computing, since artificial intelligence can potentially provide the means to achieve the properties of the «intelligent environment» of the edge-cloud continuum. The article analyzes the use of artificial intelligence in the frontier*

area of the Internet of Things. As a result of the analysis, the main advantages and problems related to the frontier area of the Internet of Things and approaches to their solution based on the use of artificial intelligence methods were determined. Edge computing and edge artificial intelligence have different applications, but the most important factor is improving the quality of service for Internet of Things devices.
Keywords: *edge computing, Internet of Things, artificial intelligence, security of edge computing.*

DOI: 10.34121/1028-9763-2024-3-4-50-68

1. Вступ

Інтернет речей (англ. Internet of Things, IoT) став потужною рушійною силою глобальної технологічної революції та промислової трансформації. Стрімкий розвиток IoT створює підґрунтя для побудови інтернету всього (англ. Internet of Everything, IoE). В епоху повного підключення сплеск кількості підключень пристроїв і обсягу даних IoT стимулює розвиток технологій граничних обчислень. Конвергенція архітектур граничних обчислень (Edge Computing) та IoT забезпечують обробку даних, зберігання та застосунки безпосередньо у пристроях IoT. Такі обчислення вирішують питання «останньої милі» комунікацій в IoT і реалізують інтелектуальне підключення та ефективне керування пристроями IoT.

Інтелектуалізація граничної області IoT забезпечує підвищення ефективності засобів ситуаційної обізнаності та управління в різних сферах застосування: в охороні здоров'я, промисловості, на транспорті та ін. Функціонування систем ситуаційного управління засноване на отриманні точної та оперативної інформації про середовище, в якому здійснюється ситуаційне управління. Конвергенція систем ситуаційного управління з засобами IoT теж забезпечує отримання точної та оперативної інформації про середовище, в якому здійснюється ситуаційне управління [1].

Інтелектуальні граничні обчислення (англ. Intelligent Edge Computing, IntEC) реалізуються шляхом розгортання алгоритмів і моделей штучного інтелекту (англ. Artificial intelligence, AI) безпосередньо на локальних граничних пристроях, таких як датчики або пристрої IoT, що забезпечує обробку та аналіз даних у реальному часі без постійної залежності від хмарної інфраструктури.

Інтелектуалізовані граничні пристрої (англ. Intelligent Edge Devices, IntED) поєднують граничні обчислення та AI для виконання завдань інтелектуальної обробки інформації безпосередньо у місці отримання інформації або керування об'єктами. Інтелектуальні граничні обчислення передбачають зберігання інформації поблизу IntED, а обробку цієї інформації з використанням AI здійснювати безпосередньо на межі мережі з підключенням до інтернету або без нього. Це забезпечує обробку даних за мілісекунди і реалізацію зворотного зв'язку у реальному часі.

Безпілотні транспортні засоби, мобільні персональні пристрої, камери безпеки та розумна побутова техніка є прикладами технологій, які використовують передові можливості AI для взаємодії з користувачами в режимі реального часу.

Використання IntED або пристроїв Edge AI дозволяє оптимізувати робочі процеси, автоматизувати бізнес-процеси і відкрити нові можливості для інновацій, одночасно вирішуючи такі проблеми, як прихованість, безпека та зниження витрат.

За участю 48 партнерів із 10 європейських країн створено консорціум EdgeAI [2], метою діяльності якого є мобілізація найкращих європейських спроможностей промисловості, досліджень і академічних кіл у сфері граничного AI, напівпровідників, мікроелектроніки, вбудованих систем, розробки програмного забезпечення AI та промислових OEM-виробників. Усі партнери вносять значний вклад у сферу досліджень і розробок в Європі, що забезпечує стабільну співпрацю відповідно до європейської стратегії, цінностей і пріоритетів як протягом, так і після завершення проєкту. Разом консорціум збирає всю необхідну базу та критичну масу, щоб закласти основи для впровадження AI у граничній обла-

сті та діє як орієнтир для того, щоб вивести Європу на передові позиції в цій галузі в усьому світі.

EdgeAI є ключовою ініціативою для європейської цифровізації на основі інтелектуальних рішень для граничних обчислень. EdgeAI розробляє нові електронні компоненти та системи, архітектури обробки, підключення, програмне забезпечення, алгоритми та проміжне програмне забезпечення за допомогою поєднання мікроелектроніки, AI, вбудованих систем і граничних обчислень. Він демонструє застосовність розроблених підходів апаратного/програмного забезпечення/алгоритмів граничного AI до різних вертикальних рішень, враховуючи вимоги безпеки, надійності та енергоефективності, властиві різним сценаріям використання. Проєкт охоплює ключові стратегічні сфери застосування, такі як цифрова економіка, енергетика, агропродовольчий сектор і виробництво напоїв, мобільність і цифрове суспільство.

Метою статті є аналіз можливостей інтелектуалізації засобів граничних обчислень у різних сферах діяльності.

2. Особливості засобів граничних обчислень

Граничні обчислення значно спрощують обробку величезних обсягів термінальних даних у хмарі. Дані, зібрані граничними пристроями, безпосередньо аналізуються та обробляються локально на межі мережі в режимі реального часу, без необхідності завантажувати їх у хмару для обробки. Граничні обчислення відповідають ключовим вимогам цифровізації галузі щодо гнучкого підключення та оптимізації даних у реальному часі. Поєднання технологій граничних обчислень та IoT породжує граничний IoT. Гранична область IoT представляє архітектуру обчислень, що інтегруються у систему IoT за допомогою граничного обчислювального шлюзу. Граничний обчислювальний шлюз, який об'єднує мережеві та обчислювальні ресурси, сховища даних та застосунки, розгортається на межі мережі поблизу пристроїв або джерел даних, щоб він міг надавати сервіси керування пристроями та управління системою IoT у граничних вузлах мережі. Отже, граничні обчислення IoT вирішують проблему «останньої милі» галузевих комунікацій IoT і впроваджують розумне підключення та ефективне керування пристроями IoT.

Переваги граничних обчислень

Граничні обчислення надають низку переваг порівняно з централізованими, віддаленими обчисленнями. Такими перевагами є:

1. Швидкість та оперативність. Обробка інформації в місці її отримання прискорює отримання результатів та забезпечує оперативне їх використання.

2. Зниження мережевого трафіка за рахунок зменшення об'ємів інформації, що потребують використання мережевих ресурсів для обміну даними.

3. Безпека даних. Зменшуються ризики втрат і перехоплення інформації під час передачі та проведення DDoS атак.

4. Локалізація інформації задовольняє вимоги її зберігання у місці отримання.

5. Результативність інформації забезпечує стійкість та відновлюваність даних за рахунок розподіленого зберігання та обробки у вузлах граничних обчислень.

6. Персоніфікація послуг за рахунок урахування локальних потреб споживача цих послуг.

7. Покращений моніторинг за рахунок наближення безпосередньо до об'єкта моніторингу.

Ці переваги граничних обчислень надають можливості реалізувати системи IoT для розумних міст, виробництва, сфери споживання тощо. Виробники використовують передові інновації в датчиках IoT і для покращення можливостей прогнозованого обслуговування, відстеження запасів і моніторингу виробничого обладнання. Граничні пристрої IoT

можна використовувати для відстеження запасів у роздрібних торговельних середовищах, а також для оптимізації логістики та запасів. У інтелектуальних застосунках граничні обчислення можуть бути ключовим інструментом, який допоможе організувати та керувати величезною кількістю інформації, отриманої від оточуючого середовища з метою забезпечення сталого розвитку.

Проблеми граничних обчислень

У розподілених гетерогенних середовищах, таких як IoT, традиційно обмеженим є перелік постачальників апаратного забезпечення та постачальників послуг, що забезпечують контрольоване наскрізне розгортання комунікаційної інфраструктури. Трансформація розподіленої архітектури з урахуванням різних типів граничних пристроїв пов'язана з впровадженням нових бізнес-моделей і моделей надання послуг (рис. 1). Продукти та послуги від різних постачальників інфраструктури та програмного забезпечення потребують налагодження узгодженої взаємодії між різними мережевими функціями та безперебійного масштабування від граничної до хмарної інфраструктури. Це породжує різні проблеми при розгортанні, масштабуванні та керуванні граничними обчисленнями. Наведемо деякі проблеми, пов'язані з використанням граничних обчислень.

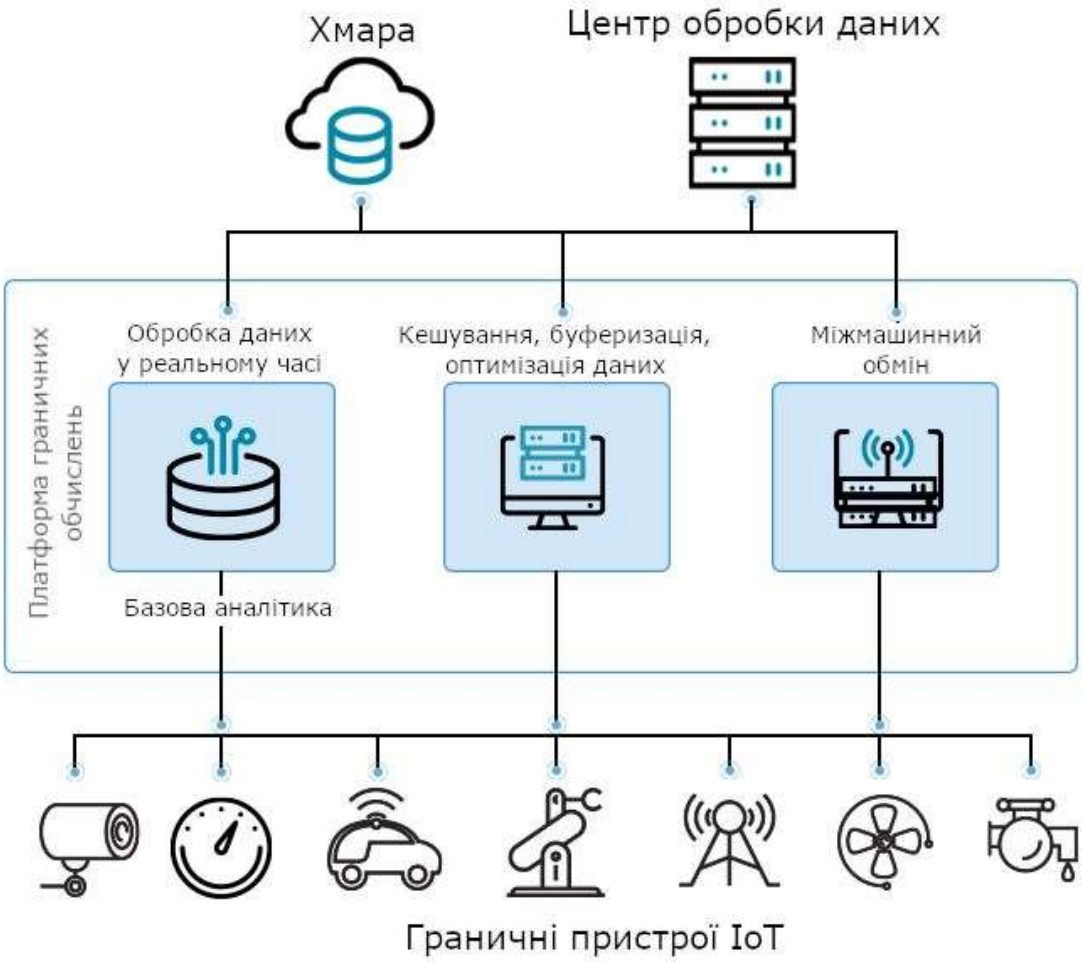

Рисунок 1 – Архітектура граничних обчислень

Витрати на обслуговування. Збільшення витрат на обслуговування обумовлене кількістю граничних пристроїв та можливим їх розподілом на великій відстані один від одного.

Пропуск даних. Граничні обчислення скорочують обсяги первинної інформації, яка передається для подальшої обробки у хмару та бізнес-застосунки. Первинна інформація може знадобитись для додаткової обробки або уточнення результатів на інших рівнях. Тому потрібно забезпечити доступ за потребою до цієї інформації у місцях її зберігання у граничній області.

Безпека граничного обладнання. Необхідно захищати розподілену мережу пристроїв. Забезпечення адекватного рівня захисту кожного пристрою може обумовити збільшення витрат на обслуговування.

Вплив середовища. Граничні пристрої часто розташовуються в місцях прямого впливу на них оточуючого середовища. Залежно від конкретного випадку використання може знадобитися спроєктувати граничні пристрої таким способом, щоб вони могли витримати екстремальні електромагнітні, механічні або біологічні впливи.

Відсутність стандартів. Конвергенція різних технологій ускладнює стандартизацію у граничній області і, як наслідок, досягнення функціональної сумісності.

Інфраструктурна підтримка. Необхідність забезпечення електроенергією та комунікаціями вимагає додаткових витрат.

Масштабування. Складність масштабування обумовлена проблемами зростання витрат на обслуговування, безпеку, інфраструктуру.

Граничні обчислення забезпечують обробку даних безпосередньо в місцях їх отримання або на локальних граничних серверах. Наближеність до джерел даних надає значні переваги в оперативності отримання обробленої інформації від об'єктів моніторингу і керування, включаючи швидку статистику, зменшення часу реакції та кращу пропускну здатність. Ці переваги є особливо важливими при побудові систем IoT. Зростання масштабів та збільшення обчислювальної потужності пристроїв IoT обумовлюють зростання обсягів даних для обробки і обміну в розподіленому середовищі з великою кількістю підключених мобільних пристроїв. Безпрецедентний масштаб і складність даних, які створюються підключеними пристроями, висувають високі вимоги стосовно спроможностей мережевої інфраструктури та її ефективного використання.

Надсилання даних, створених пристроєм, до централізованого центру обробки даних або до хмари спричиняє проблеми з пропускною здатністю та затримкою. Граничні обчислення пропонують більш ефективну альтернативу шляхом обробки та аналізу даних ближче до точки їх створення. Оскільки дані не передаються через мережу у хмару або центр обробки даних для обробки, затримка зменшується. Граничні обчислення і мобільні граничні обчислення в сучасних стільникових мережах забезпечують швидший і повніший аналіз даних, забезпечуючи більш глибоке розуміння, більш швидший час реакції та покращення підтримки користувачів.

Для збільшення спроможностей граничних обчислень, зберігання та маршрутизації даних у мережі інтернет використовуються туманні обчислення або туманні мережі. Національний інститут стандартів і технологій у березні 2018 року опублікував визначення туманних обчислень, взявши більшу частину комерційної термінології Cisco як спеціальну публікацію NIST 500-325. Концептуальна модель туманних обчислень, яка визначає туманні обчислення як горизонтальну, фізичну або віртуальну парадигму ресурсів, що знаходиться між розумними граничними пристроями та традиційними хмарними обчисленнями або центрами обробки даних [3]. Ця парадигма підтримує вертикально ізольовані програми, чутливі до затримки, забезпечуючи повсюдне, масштабоване, багаторівневе, об'єднане, розподілене обчислення, зберігання та підключення до мережі. Отже, туманні обчислення найбільше відрізняються відстанню від краю. У теоретичній моделі туманних обчислень вузли туманних обчислень фізично та функціонально працюють між граничними вузлами та централізованою хмарою [4]. Значна частина термінології не визначена,

включаючи ключові архітектурні терміни, такі як «розумний», відмінність між туманними обчисленнями та граничними обчисленнями не є чітко визначеною.

Поєднання граничних, туманних та хмарних обчислень

У той час, як граничні обчислення називають місцем, де створюються екземпляри сервісів, туманні обчислення передбачають розподілені на пристроях і системах або поблизу них комунікації, обчислення, ресурси зберігання та сервіси, якими керують кінцеві користувачі. Туманне обчислення є полегшеним проміжним рівнем обчислювальної потужності. Туманні обчислення часто служать доповненням до хмарних обчислень. Туманні обчислення є більш енергоефективними, ніж хмарні.

Ініціатива EUCloudEdgeIoT.eu [5] спрямована на реалізацію розуміння та розвитку континууму Cloud, Edge та IoT (CEI) шляхом сприяння співпраці між широким спектром дослідницьких проєктів, розробниками та постачальниками, бізнес-користувачами та потенційними користувачами цієї нової технології. Edge to Cloud (IoT–Edge–Cloud або Cloud, Edge and IoT, CEI) Continuum — це концепція, яка об'єднує хмарні обчислення та стільникові мережі і набирає популярності завдяки своєму потенціалу забезпечувати безперебійну роботу користувача та вирішувати проблеми керування складними багатодоменними мережами, що включають масиви пристроїв IoT (рис. 2). Мета підходу IoT–Edge–Cloud Continuum полягає в тому, щоб розподілити обчислення та навантаження даних між декількома типами пристроїв, використовуючи переваги кожного, наприклад, близькість до джерела даних, доступ до даних або обчислювальну потужність, одночасно зменшуючи слабкі сторони [6]. Включення інтелекту в Edge to Cloud Continuum може ще більше розширити його можливості, пропонуючи такі переваги, як зменшення затримки, покращену масштабованість, покращене використання ресурсів і покращене усвідомлення контексту.

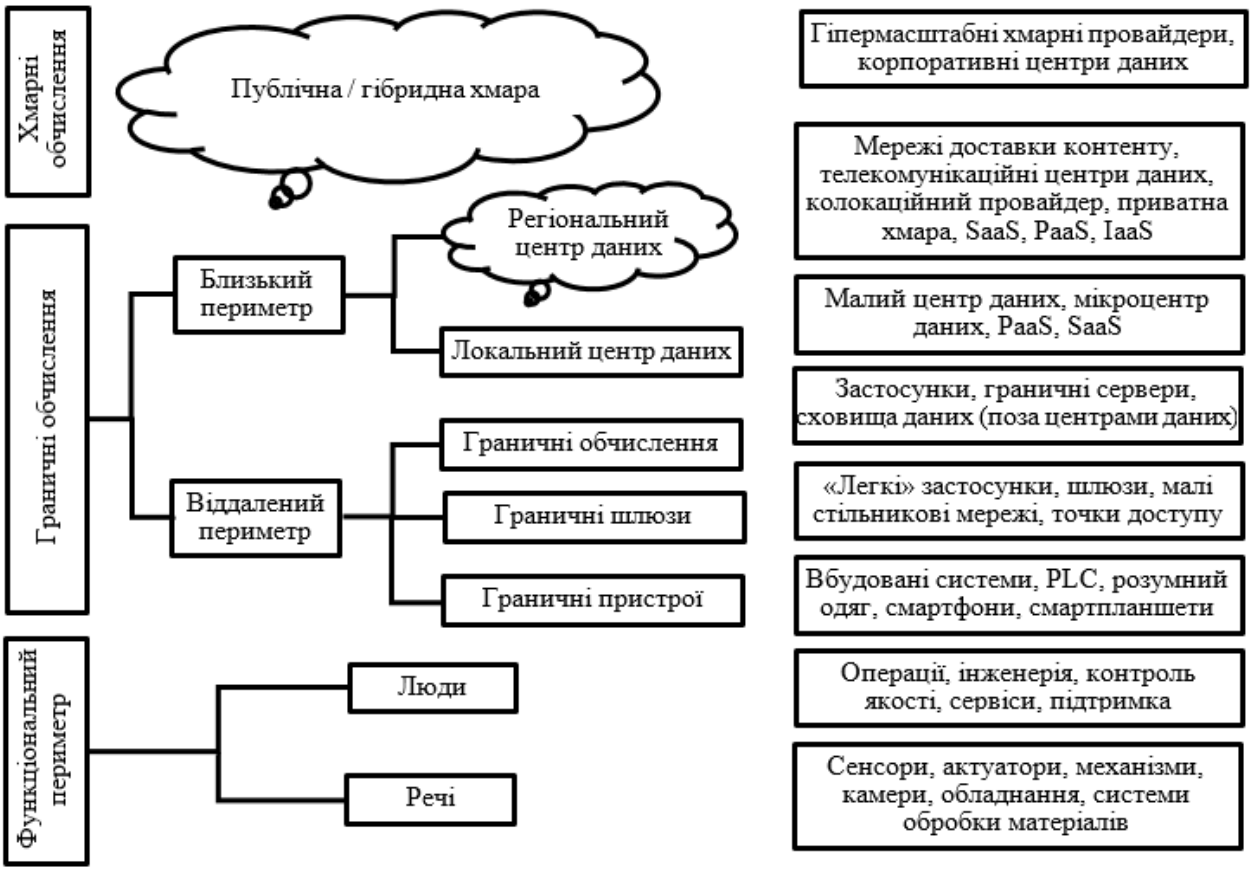

Рисунок 2 — Континуум CEI (Edge to Cloud Continuum)

3. Моделі інтелектуалізації граничних обчислень

Технологічний прогрес і зростаюча комерціалізація граничних/туманних обчислень обумовили інтенсифікацію досліджень, пов'язаних із граничним AI (Edge AI). Інтелектуалізація граничних обчислень означає розгортання алгоритмів і моделей AI безпосередньо на локальних граничних пристроях IoT, таких як сенсори, вимірювальні перетворювачі, актуатори, що забезпечує обробку та аналіз даних у реальному часі без постійної залежності від хмарної інфраструктури. Тобто граничний AI або «AI на межі» означає поєднання граничних обчислень і AI для виконання завдань машинного навчання безпосередньо на взаємопов'язаних граничних пристроях. Граничні обчислення дозволяють зберігати дані поблизу пристрою, а алгоритми AI дозволяють обробляти дані безпосередньо на краю мережі з підключенням до інтернету або без нього. Це полегшує обробку даних за мілісекунди, забезпечуючи зворотний зв'язок у реальному часі.

Граничний AI має кілька рівнів залежно від ролей граничних вузлів (наприклад, граничних вузлів мережі, пристроїв користувача) у створенні функцій AI. На базовому рівні крайові вузли використовують функції AI та машинного навчання (ML), створені десь в іншому місці (наприклад, у хмарі), але не беруть участі у створенні цих функцій. На верхньому рівні Edge learn вивчає свої локальні дані (тобто граничне навчання), щоб допомогти створити моделі AI/ML не лише для себе, але й для інших мережевих об'єктів і програм користувача.

Оскільки великий обсяг даних створюється на межах мережі й відправлення їх у хмару є проблематичним, граничне (крайове, граничне) навчання стає необхідним і важливим для навчання майбутніх моделей ML для розширених мереж, обчислювальних інфраструктур і програм користувача. Проте навчання на межах стикається з фундаментальними проблемами, які існуючі методи машинного навчання не можуть адекватно вирішити. Такі виклики включають обмеження ресурсів, неідентичний або незалежний (англ. non-identical or non-independent, non-IID) розподіл даних, вимоги конфіденційності даних, обмеження зв'язку, а також вразливості безпеки.

Моделі навчання, алгоритми AI та комп'ютерні програми на їх основі постійно вдосконалюються, розширюючи можливості прогнозування, розуміння складної інформації та автоматичного прийняття раціональних рішень. Динамічно розвиваються такі методи машинного навчання, як:

- навчання з підкріпленням, яке допомагає комп'ютерам робити кращий вибір один за одним, не потребуючи попередніх прикладів;
- федеральне (інтегроване) навчання, яке дозволяє комп'ютерам навчатися разом, не надсилаючи всі свої дані в одне місце;
- генеративний AI для створення нових даних, коли реальних даних недостатньо;
- квантове машинне навчання, завдяки якому комп'ютери можуть набагато швидше обробляти інформацію;
- перенесене навчання.

Впровадження цих методів роблять граничні пристрої IoT інтелектуальними у своїх діях. Безпілотні автомобілі, мобільні пристрої, камери безпеки та розумна побутова техніка є прикладами технологій, які використовують передові можливості AI, щоб надавати користувачам інформацію в режимі реального часу, коли у цьому виникає потреба.

Граничний AI (Edge AI) стає все популярнішим, оскільки в індустрії знаходять нові способи використання його можливостей для оптимізації робочих процесів, автоматизації бізнес-процесів і відкриття нових можливостей для інновацій, одночасно вирішуючи такі проблеми, як затримка, безпека та зниження витрат на основі конвергенції технологій [7–11].

Edge AI означає розгортання алгоритмів і моделей AI безпосередньо на граничних пристроях, таких як смартфони, пристрої IoT, датчики та інші вбудовані системи, замість

того, щоб покладатися виключно на централізовані хмарні сервери. Узагальнена модель конвергенції AI з граничними обчисленнями наведена на рис. 3.

Традиційно завдання AI, які вимагали складних обчислень, виконувалися у віддалених центрах обробки даних або хмарних серверах через їх високі вимоги до обчислень і потребу у значних ресурсах. Однак такий підхід часто створює затримку, споживає значну пропускну здатність мережі та викликає занепокоєння щодо конфіденційності та безпеки даних, особливо під час роботи з конфіденційною інформацією. Edge AI вирішує ці проблеми, надаючи можливості AI безпосередньо на граничні пристрої, де генеруються та споживаються дані. Це означає, що обробка даних, аналіз і прийняття рішень відбуваються локально на самому пристрої, без необхідності передавати дані на центральний сервер для обробки. Цей підхід має кілька переваг.

Зменшена затримка: обробляючи дані локально, граничний AI скорочує час, потрібний для аналізу даних і виконання дій, покращуючи час відповіді для додатків у реальному часі.

Рисунок 3 — Модель конвергенції граничних обчислень та AI

Покращена конфіденційність і безпека: оскільки дані залишаються на пристрої і не передаються через мережу, Edge AI може підвищити конфіденційність і безпеку, мінімізуючи ризик витоку даних або несанкціонованого доступу.

Ефективність пропускної здатності: Edge AI зменшує потребу в передачі великих обсягів даних на централізовані сервери, що приводить до зниження вимог до пропускної здатності та потенційного зниження витрат, пов’язаних із передачею даних.

Функціональність в автономному режимі: Edge AI дозволяє пристроям виконувати завдання AI, навіть якщо вони не підключені до інтернету, покращуючи функціональність у середовищах з обмеженим або періодичним підключенням.

Масштабованість: Edge AI розподіляє обчислювальне навантаження між декількома пристроями, дозволяючи масштабовані програми AI без надмірного навантаження на централізовані сервери.

Поширені програми граничного AI включають відеоаналітику в реальному часі, прогнозне технічне обслуговування промислового обладнання, розумні домашні пристрої, автономні транспортні засоби та переносні монітори стану здоров’я тощо.

Граничні обчислення допомагають зробити зберігання даних і обчислення більш доступними для користувачів. Це досягається за допомогою виконання операцій на локальних пристроях, таких як ноутбуки, пристрої IoT або виділені граничні сервери. На граничні процеси не впливають проблеми з затримкою та пропускною спроможністю, які часто перешкоджають роботі хмарних операцій.

Edge AI поєднує граничні обчислення з AI. Це передбачає запуск алгоритмів AI на локальних пристроях із граничними обчислювальними можливостями. Edge AI не вимагає підключення та інтеграції між системами, що дозволяє користувачам обробляти дані на пристрої в режимі реального часу.

Більшість процесів AI зараз виконується у хмарних центрах, оскільки вони вимагають значних обчислювальних потужностей. Недоліком є те, що проблеми з підключенням або мережею можуть призвести до простою або значного сповільнення служби. Edge AI усуває ці проблеми, роблячи процеси AI невід'ємною частиною граничних обчислювальних пристроїв. Це дозволяє користувачам економити час, збираючи дані та обслуговуючи користувачів, не зв'язуючись з іншими фізичними місцями.

Федеративне навчання (часто його називають спільним навчанням) — це децентралізований підхід до навчання моделей машинного навчання. Для цього не потрібен обмін даними з клієнтських пристроїв на глобальні сервери. Натомість необроблені дані на граничних пристроях використовуються для локального навчання моделі, підвищуючи конфіденційність даних [12].

Розглянемо три різні підходи до моделей розгортання AI та міркувань, а також їхні плюси та мінуси.

AI у хмарних обчисленнях

Хмарні обчислення забезпечували високомасштабоване недороге апаратне забезпечення, яке було переконливим для AI, оскільки дозволяло організаціям швидко навчати великомасштабні моделі. Однак, незважаючи на те, що хмара дуже підходить для навчання моделі, це може бути складним для діалогових та конвеєрних міркувань — використання моделей AI для надання прогнозів у відповідь на запити користувачів.

Використання хмари для міркувань викликає кілька проблем:

1. Хмарне міркування може мати проблеми з відповідями в реальному часі, які необхідні для багатьох випадків використання AI. Це пояснюється тим, що спочатку необхідно передавати запит із граничного пристрою у хмару, а потім відповідь назад на граничний пристрій.

2. Навіть якщо сценарій використання не вимагає відповіді в реальному часі, хмарний висновок за своєю природою має високу затримку, що погіршує взаємодію з користувачем.

3. Якщо граничний пристрій не має з'єднання з інтернетом або має проблеми з підключенням, він взагалі не може виконати хмарний висновок. Навіть якщо пристрій має підключення до інтернету, він може не мати достатньої пропускної здатності для передачі відповідного обсягу даних за прийнятний проміжок часу.

Edge AI

Edge AI моделі AI працюють на граничних пристроях без затримок і підключення до інтернету. Це дає можливість приймати більш оперативні рішення та підтримувати варіанти використання в реальному часі.

Однак у Edge AI також виникає кілька проблем, оскільки моделі потрібно постійно навчати, використовуючи дані з граничних пристроїв. Отже:

1. Системі потрібно створити набір даних, передаючи дані з великої кількості граничних пристроїв у хмару. Це може бути складним і важким для досягнення мети залежно від підключення та пропускної здатності, доступної для граничних пристроїв.

2. Зберігання всіх даних у централізованому місці створює проблеми з конфіденційністю та відповідністю. Таке законодавство, як GDPR, ускладнює навчання моделей AI на основі даних кінцевих користувачів, а централізована база даних становить загрозу безпеці.

Федеративне (інтегроване, спільне) навчання

Модель, відома як федеративне (інтегроване) навчання, може вирішити проблеми хмарних обчислень і граничного AI. Цей шаблон працює так:

1. Моделі AI навчаються на граничних пристроях, використовуючи їх локальні дані.

2. Оновлення моделі надсилаються на центральний сервер без необхідності надсилати фактичні дані пристрою, що вирішує багато проблем із конфіденційністю та безпекою.

3. Оновлення моделі об'єднуються в консолідовану модель, а оновлена модель повертається на клієнтські пристрої.

Edge AI з Run:AI

Edge AI поєднує граничні обчислення з AI. Це передбачає запуск алгоритмів AI на локальних пристроях із граничними обчислювальними можливостями. Edge AI не вимагає підключення та інтеграції між системами, що дозволяє користувачам обробляти дані на пристрої в режимі реального часу. Kubernetes платформа, на якій базується планувальник Run:AI [13], має полегшену версію під назвою K3s, призначену для обчислювальних середовищ з обмеженими ресурсами, таких як Edge AI. Run:AI автоматизує та оптимізує керування ресурсами та оркестрування робочого навантаження для інфраструктури машинного навчання. За допомогою Run:AI можна запускати більше робочих процесів на серверах з обмеженими ресурсами.

Використання Run:AI забезпечує такі можливості:

• Розширена видимість. Створення ефективного конвеєра спільного використання ресурсів шляхом об'єднання обчислювальних ресурсів GPU.

• Усунення вузьких місць. Виконувати робочі навантаження можна на частинах GPU та ефективніше керувати пріоритетами.

Вищий рівень контролю. Run:AI дозволяє динамічно змінювати розподіл ресурсів, гарантуючи, що кожне завдання отримує необхідні ресурси в будь-який момент часу.

Run:AI спрощує конвеєри інфраструктури машинного навчання, допомагаючи дослідникам даних підвищити продуктивність і якість своїх моделей глибокого навчання.

4. Архітектури інтелектуалізованих граничних обчислень

Перенесення обчислювальних навантажень у граничну область IoT, а не у хмару обумовлено низкою факторів. Насамперед, у граничній області створюються дані. Це місце, де локалізовані операційні технології (OT) користувачів і виробників. Надсилання даних у хмару потребує додаткових витрат. Окрім вартості надсилання та зберігання даних, слід враховувати вартість використання обчислювальних ресурсів у хмарі. Сьогодні хмарні послуги користуються великим попитом, і вартість обчислювальних ресурсів у хмарі може бути значною. З іншого боку, якщо обробка даних виконується у граничній області на апаратному забезпеченні, яким володіє користувач, єдиною ціною є вартість апаратного забезпечення, витрати на його експлуатацію та обслуговування, які можна додатково зменшити шляхом використання відповідного промислового обладнання [14].

Крім вартості, затримка є ще одним важливим моментом. Це може бути проблематично та обмежувати залежно від варіанта використання, оскільки швидкі рішення часто потрібні на виробництві або у граничній області IoT. Якщо, наприклад, щойно вироблені з'єднувачі спускаються з виробничої лінії з високою швидкістю, вкрай важливо, щоб рішення про те, який з'єднувач пройшов перевірку якості, а який не пройшов, було прийнято майже миттєво. Крім того, якщо ваша виробнича лінія покладається на рішення, прийняті виключно у хмарі, втрата з'єднання, швидше за все, призведе до простою виробничої лінії, втрачаючи гроші, час і цінну пропускну здатність.

Edge AI — це конвергенція багатьох технологій, включаючи AI, IoT, граничні обчислення та вбудовані системи, кожна з яких відіграє вирішальну роль у забезпеченні інтелектуальної обробки та прийняття рішень на межі мережі. Edge AI передбачає використання вбудованих алгоритмів для моніторингу активності віддаленої системи, а також обробки даних, зібраних такими пристроями, як датчики та інші трекери неструктурованих даних, включаючи температуру, мову, обличчя, рух, зображення, наближення та інші аналогові дані.

Ці граничні системи можуть приймати різні форми, включаючи датчики, смартфони, пристрої IoT, дрони, камери та навіть транспортні засоби й розумні пристрої. Дані, зібрані з цих систем, служать вхідними даними для алгоритмів граничного AI, надаючи цінну інформацію про стан системи або її оточення, дозволяючи системам граничного AI швидко реагувати на зміни або аномалії та розуміти середовище, в якому вони працюють. Ці граничні додатки AI були б непрактичними або навіть неможливими для роботи в централізованій хмарі або середовищі корпоративного центру обробки даних через проблеми, пов'язані з вартістю, затримкою, пропускною здатністю, безпекою та конфіденційністю.

Одним із напрямів еволюції програмного забезпечення, пов'язаного з граничними обчисленнями, є застосування методів і алгоритмів AI як безпосередньо для обробки інформації, так і для адаптивного програмно керованого розгортання мережевої інфраструктури. Вже існують системи, які реалізують принаймні деякі з рішень. Наприклад, розгортання SD-WAN [15], uCPE [16] або vRAN [17].

Граничний інтелект, як і багато інших функцій CEI, значною мірою можливий завдяки прогресу в апаратному забезпеченні граничних пристроїв. Останніми роками активний розвиток нових типів спеціалізованих інтегральних схем Edge AI, розроблених спеціально для роботи у граничній області IoT, зокрема, блоків обробки зору (наприклад, Intel Movidius VPUs2) [18], блоків обробки тензорів (наприклад, Google Edge TPU) [19], блоків нейронної обробки (наприклад, процесор машинного навчання Arm Ethos U554) [20] та ін.

У машинному навчанні є два основні обчислювальні процеси: навчання та висновок. Ці процеси потенційно можуть бути розташовані на будь-якому рівні в ієрархії «пристрій-хмара» або навіть на комбінації рівнів. Різні автори виділяють шість або сім рівнів розміщення ML в ECC.

Ці рівні включають навчання та висновок у хмарі, навчання у хмарі з сумісним висновком на межі хмари, спільне навчання та спільний висновок на межі хмари, навчання та логічний висновок на краю, навчання та висновок на пристрої та проміжні варіанти між ними.

Крім тренування та визначення краю, граничний інтелект також включає два інші процеси: кешування краю та розвантаження краю. Граничне кешування передбачає зберігання даних, згенерованих на пристроях з невеликими можливостями зберігання та обробки, на більш потужних вузлах, таких як хмарні програми, розташовані в автомобілях, або граничні сервери. Потім збережені дані використовуються як вхідні дані для процесів навчання та висновків.

Окрім зберігання необроблених даних (наприклад, відеоканали з камер), кешування меж також використовується для зберігання результатів операцій виведення (наприклад, результатів розпізнавання об'єктів у відеопотоках). Це робиться для того, щоб зменшити використання обчислювальних ресурсів для висновків і зробити ці результати доступними для граничних пристроїв, які не мають власних можливостей висновків.

На рис. 4 показано взаємозв'язок між процесами, пов'язаними з граничним інтелектом, і пристроями, які з ним взаємодіють.

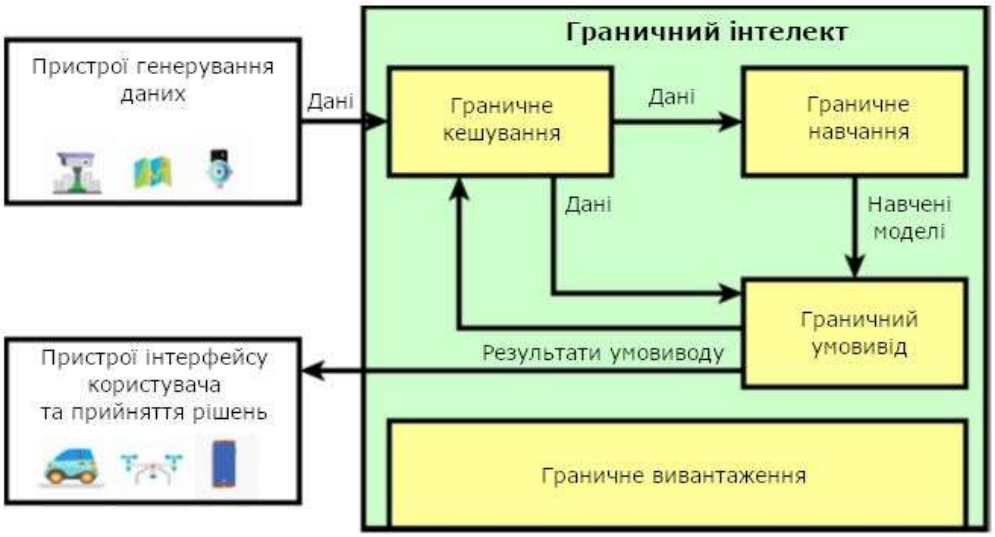

Рисунок 4 — Архітектура інтелектуальних граничних обчислень

Li W. та ін. [21] запропонували діаграму розподілу для машинного навчання на Edge (див. рис. 5). Тут DL є частиною гілки навчання і поділяється на два методи: федеративне навчання (FL) і розщеплене навчання (SL).

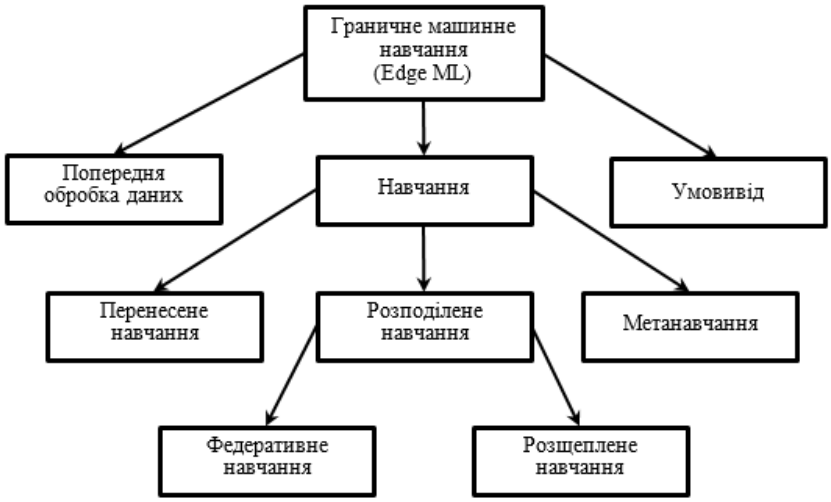

Рисунок 5 — Ієрархія методів машинного навчання у граничній області

Швидкий розвиток і поширення технологій AI, зокрема глибокого навчання, є ще однією сучасною тенденцією, яка супроводжує поточний зсув у бік граничних обчислень. Зближення цих тенденцій, імовірно, продовжуватиметься, оскільки AI потенційно може забезпечити засоби досягнення деяких із згаданих вище властивостей «розумного середовища» ECC. Зокрема, динамічна ситуаційна співпраця в ECC може бути досягнута шляхом посилення компонентів програмного забезпечення властивостями агентів із когнітивними можливостями, такими як здатність до навчання та проактивного визначення курсу дій щодо самоадаптації.

Крім того, високорозподілена архітектура ECC надає нові можливості для програм машинного навчання. Величезні обсяги даних, що генеруються кінцевими пристроями, можуть бути оброблені на місці (за допомогою сумісного навчання та сумісного висновку граничної хмари пристрою), не перевантажуючи мережі надмірною передачею даних.

Отже, ми можемо розраховувати на більш широке використання федеративного навчання, спільного навчання та спільного висновку між пристроєм і хмарою.

5. Приклади застосування інтелектуалізованих граничних обчислень

Edge AI — це інноваційна технологія, яка впроваджується у численні галузі та пропонує перспективні рішення. Інтелектуальні граничні обчислення спираються на пристрої, здатні оцінювати дані, швидко приймати рішення та безпечно з'єднуватися в режимі реального часу, при цьому значно знижуючи витрати на електроенергію та використання. Завдяки численним перевагам, таким як підвищена ефективність, заходи безпеки та низьке енергоспоживання, Edge AI здійснює величезний вплив на наше життя. Розглянемо деякі приклади застосування інтелектуальних граничних обчислень. Edge AI охоплює широкий спектр варіантів використання. Розглянемо деякі з них.

Промисловість

Завдяки розгортанню алгоритмів AI безпосередньо на виробничому обладнанні та датчиках граничні пристрої можуть контролювати стан обладнання, виявляти дефекти та оптимізувати виробничі процеси, не покладаючись на централізовані сервери. Для Idustry 4.0 дані збираються та агрегуються у граничних вузлах мережі промислового інтернету речей (англ. Industrial Internet of Things, IIoT). Наприклад, туманний вузол граничної області здатний збирати та агрегувати дані з мережі IIoT, а потім використовувати ці дані для розгортання різних граничних рішень AI. Рішення для Edge AI передбачають розгортання глибоких нейронних мереж (DNN) зі спеціальними стратегіями для пом'якшення проблем, пов'язаних із великою кількістю пристроїв. Зокрема, ці проблеми можуть бути вирішені з використанням активного навчання (англ. active learning, AL) для вирішення проблеми немаркованих даних, перенесеного навчання (англ. transfer learning, TL) для забезпечення попередньо навченої моделі та федеративне навчання (англ. federated learning, FL) для побудови глобальної моделі із забезпеченням конфіденційності. Спочатку аналізуються, моделюються та порівнюються всі три граничні рішення AI. Далі розглядається їхня комбінація, яка називається федеративним активним перенесеним навчанням (англ. federated active transfer learning, FATL). Таке комплексне рішення могло б вирішити всі проблеми, пов'язані із застосуванням граничного AI в мережах IIoT.

Сільське господарство

Edge AI підтримує точне землеробство, аналізуючи дані з датчиків, безпілотних літальних апаратів і супутникових зображень для моніторингу здоров'я врожаю, оптимізації зрошення, виявлення зараження шкідниками та аналізу умов навколишнього середовища в реальному часі. Agriculture 4.0 використовує IoT, кіберфізичні системи, інтелектуальних агентів та аналіз великих даних у безперебійній аграрній екосистемі, щоб забезпечити продовольчу безпеку та безпеку за менших витрат. Останніми роками граничні обчислення стають все більш популярною темою та дозволяють розгортати багато програм IoT у різних галузях. Граничні обчислення надають можливості для підвищення ефективності та зручності використання сільськогосподарських систем шляхом полегшення обробки даних на межі за допомогою методів AI та уможливлення проактивного прийняття рішень із використанням великих даних. Нові сфери AI, граничних обчислень та IoT містять великий потенціал для розширення можливостей фермерів використовувати кращі ресурси в щоденному виконанні ефективних і стійких процесів. Ці передові концепції реалізуються завдяки конвергентним технологіям і системам, у тому числі бездротовим технологіям, таким як робототехніка та сільськогосподарські машини, хмарні обчислення й аналітика великих даних, мітки RFID/NFC, сенсорні мережі, аналітика великих даних і високопродуктивні обчислення.

Конвергенція сільського господарства та AI (Agri-AI) стала критично важливим рішенням, яке пропонує потенціал для підвищення врожайності, забезпечення безпеки харчових продуктів і революції в сільськогосподарських практиках. Впровадження AI в сільське господарство є значним кроком вперед, обіцяючи покращену точність і ефективність у різних сільськогосподарських процесах.

Безперервний технологічний прогрес і зростаючий попит на методи сталого ведення сільського господарства підкреслює вирішальну роль машин, керованих AI, у підвищенні продуктивності сільського господарства шляхом автоматизації трудомістких завдань, забезпечення точності посіву та збирання врожаю й використання ресурсів.

Edge AI переносить потужність AI безпосередньо в сільське господарство. Це впровадження на місці полегшує прийняття рішень у реальному часі, мінімізуючи затримку та значно покращуючи загальне управління фермою. Завдяки технологіям Edge AI, наприклад, лінійці продуктів Aetina [22], фермери можуть аналізувати дані там, де вони генеруються, що дозволяє їм швидко приймати обґрунтовані рішення та виконувати точні сільськогосподарські методи з неперевершеною миттєвістю. Це не тільки приводить до значного підвищення продуктивності, але також і до значної економії коштів і ресурсів для фермерів.

Охорона здоров'я

Edge AI підтримує дистанційне спостереження за пацієнтами та персоналізовану медичну допомогу, аналізуючи дані з переносних пристроїв, медичних датчиків і обладнання для обробки зображень, щоб виконувати аналіз медичних даних у реальному часі та сповіщати постачальників медичних послуг про можливі проблеми зі здоров'ям. При застосуванні Edge AI у секторі охорони здоров'я слід враховувати такі критично важливі фактори, як обробка даних у реальному часі, конфіденційність і безпека.

Дистанційне спостереження за пацієнтом: Edge AI дозволяє переносним пристроям, оснащеним датчиками, контролювати життєво важливі показники, такі як частота серцевих скорочень, артеріальний тиск і рівень глюкози в крові, у режимі реального часу. Ці пристрої можуть аналізувати дані локально та сповіщати постачальників медичних послуг або пацієнтів про будь-які аномалії, дозволяючи своєчасно втручатися та проактивно керувати медичною допомогою.

Розумне місто

Розумне місто використовує інформаційно-комунікаційні технології (ІКТ) для підвищення ефективності міських послуг, оптимізації використання ресурсів і покращення загальної якості життя його мешканців. Це включає дані, отримані датчиками, автоматизовані процеси та аналіз даних, які об'єднуються, щоб допомогти міській владі приймати кращі рішення для сталого та розумного міського розвитку. Дані з датчиків і камер, розміщених по всій міській території, можуть використовувати різноманітні додатки для розумного міста, включаючи керування дорожнім рухом, моніторинг громадської безпеки, управління відходами та оптимізацію енергоспоживання.

Розумні міста спрямовані на вирішення проблем, пов'язаних із транспортом, енергетикою, охороною здоров'я та іншими ключовими аспектами міського життя за допомогою інноваційних технологічних рішень. Технологія розумного міста також пропонує орієнтовані на людей рішення, такі як покращений зв'язок у разі надзвичайних ситуацій або полегшення розмов із представниками уряду чи комунальних служб.

Автономні транспортні засоби

Edge AI дозволяє транспортним засобам аналізувати дані датчиків у режимі реального часу, щоб миттєво приймати рішення для таких завдань, як виявлення об'єктів, відстеження смуги руху та уникнення зіткнень, без постійної залежності від підключення до хмари.

6. Інтелектуалізація безпеки граничних обчислень

Незважаючи на переваги безпеки, притаманні рішенням Edge, розвиток Edge-обчислень і Edge AI породив інший набір проблем безпеки, що викликало потребу в нових підходах. Мобільні програми та додатки IoT є головними рушійними силами Edge-обчислень і, відповідно, безпеки Edge. Через зростання кількості пристроїв IoT існує попит на рішення, які можуть обробляти величезні обсяги даних, забезпечуючи при цьому безпеку. Багато компаній не наважуються надсилати конфіденційні дані про свою компанію у хмару та віддають перевагу зберігати їх у локальних сховищах, тим самим зменшуючи ризики кібербезпеки. Конвергенція IT та OT збільшує потенційну площу атаки. Важливо, щоб підприємства зміцнили свої граничні обчислювальні вузли для захисту своїх даних. Багато технологій, у тому числі SASE (Secure Access Service Edge) [23], а також Zero Trust, все частіше застосовуються для підвищення безпеки даних у граничній області.

Концепція Secure Access Service Edge (SASE) відноситься до категорії послуг і апаратного забезпечення, що використовується для посилення безпеки Edge. Термін був введений міжнародною консалтинговою та дослідницькою компанією Gartner у 2019 році. Gartner пояснює, що SASE — це зростаюча сила, яка може ефективно поєднувати широкі функції безпеки мережі та комплексні можливості WAN. У тому ж році, коли був введений термін, Gartner передбачив, що до 2024 року майже 50 % підприємств приймуть стратегії SASE.

Гранична служба безпечного доступу (SASE) — це технологія, яка використовується для доставки сервісів глобальної мережі (англ. wide area network, WAN) і засобів керування безпекою як служби хмарних обчислень безпосередньо до джерела з'єднання (користувача, пристрою, пристрою IoT або розташування граничних обчислень), а не до центру обробки даних. SASE використовує технології хмарних і граничних обчислень, щоб зменшити затримку, яка є результатом передачі всього трафіка глобальної мережі на великі відстані до одного або кількох корпоративних центрів обробки даних через збільшення кількості розсіяних користувачів та їхніх програм за межами приміщення. Це також допомагає організаціям підтримувати розсіяних користувачів.

Безпека SASE базується на цифровій ідентифікації, контексті в реальному часі, політиках компанії та нормативних вимогах, а не на пристрої безпеки, як-от брандмауер. Цифрова ідентифікація може бути прикріплена до будь-чого, від людини до пристрою, хмарної служби, прикладного програмного забезпечення, системи IoT або будь-якої обчислювальної системи.

Гранична безпека є складовою корпоративної безпеки для корпоративних ресурсів, які знаходяться за межами периметрів безпеки централізованого центру обробки даних. Цей тип безпеки захищає програми та користувачів, розташованих у граничній області мережі компанії, де дані вразливі до загроз безпеки.

Найпоширеніші ризики безпеки граничної області пов'язані з такими факторами [24].

1. Взаємодія з мережею може бути ризикованою для пристроїв із підтримкою Edge. Однією з важливих проблем, пов'язаних з Edge-обчисленнями, є те, як пристрої з підтримкою Edge повинні взаємодіяти з мережею, але ця мережа може їх не захищати. Під час використання Edge computing і Edge AI технологія в середині пристрою сама керує конфіденційними даними, і пристрій сам по собі може бути не таким безпечним, як мережа. Інша проблема полягає в тому, що мережа може бути під загрозою, якщо скомпрометована система Edge computing або Edge AI. Ці ризики можна зменшити, поєднавши логічні та фізичні заходи безпеки. Логічні заходи безпеки включають встановлення надійних засобів авторизації, автентифікації та шифрування даних. Навпаки, фізичні заходи безпеки включають гарантію того, що пристрій добре захищено, і контроль за тим, щоб фізичний доступ мали лише авторизовані особи.

2. Граничні пристрої викликають проблеми з кібербезпекою. Коли дані компанії зберігаються та до них отримують доступ пристрої на межі, ризик кібербезпеки може бути вищим. Для пристроїв Edge, створених без вбудованого захисту, зловмисне програмне забезпечення стає справжньою проблемою. Що ще гірше, багато готових пристроїв Edge або не мають необхідних функцій безпеки, або оновлюються не так регулярно, як це повинно бути. Моделі машинного навчання (англ. machine learning, ML), які забезпечують роботу систем Edge AI, також можуть бути підроблені, якщо не вжити належних заходів безпеки. Якщо злочинці знайдуть способи використовувати лазівки в технології, системи, які залежать від можливостей Edge AI (наприклад, камери спостереження), можуть стати серйозною загрозою безпеці. Блокування моделей машинного навчання шляхом розгляду їх як основного активу, який потребує захисту, буде життєво важливим для уникнення кібератак. Також важливо переконатися, що алгоритми Edge AI регулярно контролюються та оновлюються.

3. Наявність технології Edge полегшує атаку. Оскільки технологія Edge стає доступною, а багато пристроїв із підтримкою Edge відносно недорогі, майже кожен може отримати цю технологію та проаналізувати її на наявність уразливостей. Коли зловмисники мають доступ до пристроїв із підтримкою Edge AI, створити шкідливе програмне забезпечення, призначене для компрометації технології, стає набагато простіше. Через це пристрої Edge піддаються більшому ризику атаки.

Edge Security ma IoT

Граничні обчислення та Edge AI мають кілька варіантів використання, але один важливий варіант використання — підвищення якості обслуговування для пристроїв IoT. Завдяки пристроям IoT величезна кількість точок входу на Edge сприяє його вразливості. А кількість діючих пристроїв IoT продовжує зростати, погіршуючи ситуацію. Не кажучи вже про те, що багато пристроїв IoT не мають надійних функцій безпеки.

Слід визнати, що не всі організації, які виробляють пристрої IoT, належним чином захищають їх. Отже, необхідно, щоб самі мережі були захищені, щоб компенсувати неналежний захист граничного обладнання.

Організації можуть здійснювати керування, запрограмувавши спеціальні заходи безпеки на вузлах Edge, які можуть працювати як мікроцентри обробки даних із більшою обчислювальною потужністю та розширеними функціями безпеки. У цих випадках будь-який трафік, що передається від скомпрометованих пристроїв IoT, позначається як пошкоджений і не може отримати доступ до мережі.

6.1. Способи захисту граничних обчислень

Враховуючи, що для Edge AI актуальними є мережеві загрози, розглянемо деякі методи, які можна використати для пом’якшення цієї проблеми.

1. Компанії повинні запровадити стандарти граничних обчислень. Оскільки системи Edge AI все ще відносно нові, одним із найбільш значних ризиків є неадекватність стандартів для розробки та процесів Edge AI. Наявність чітких стандартів допомагає захистити безпеку пристрою та мережі, а також визначає зручний набір інструкцій, яких компанії можуть дотримуватись у майбутньому.

2. Раннє виявлення загроз допомагає захистити граничну область. Оскільки граничні обчислення не є централізованими, постачальники повинні індивідуально впроваджувати на практиці технологію виявлення загроз, яка може ефективно виявляти потенційні порушення, перш ніж вони спричинять шкоду та втрату даних.

3. Управління вразливістю є обов’язковим для Edge computing і Edge AI. Обслуговування та пошук відомих і невідомих уразливостей необхідно проводити постійно, якщо розглядати Edge computing і Edge AI.

4. Граничні рішення вимагають відмінної безпеки периметра. Безпека периметра може включати брандмауери та тунелі шифрування для захисту доступу до обчислювальних ресурсів Edge.

5. Безпека додатків має важливе значення для Edge-обчислень і Edge AI. Щоб бути в безпеці, програми, які працюють на граничних обчисленнях і пристроях Edge AI, повинні бути захищені за межами мережевого рівня.

6. Цикли виправлень можуть допомогти захистити пристрої на межі. Необхідно запровадити керування виправленнями, щоб оновлювати пристрої та зменшувати ймовірність поверхневих атак. Управління виправленнями стосується розповсюдження та застосування оновлень програмного забезпечення для виправлення помилок та інших уразливостей.

7. Шифрування є важливим компонентом безпеки Edge. Якщо компанії прагнуть запропонувати надійну безпеку на Edge, їм спочатку потрібно переконатися, що всі дані, які проходять через кінцеві точки компанії (і всі дані, що зберігаються на пристроях компанії), зашифровані.

6.2. Стратегія безпеки граничних обчислень

Пристрої Edge з підтримкою AI та Edge-обчислення пропонують великий потенціал. Але, як і всі технології, Edge computing і Edge AI також мають можливі ризики для безпеки. Під час виконання граничних обчислень необхідно попередньо виявити вразливі місця, після чого їх потрібно усунути, щоб захистити систему та заблокувати зловмисні атаки.

Загалом, доброякісна стратегія безпеки Edge повинна брати до уваги такі фактори:

– шифрування даних при обміні та зберіганні у сховищах;

– обмежений доступ до мережі та ресурсів даних;

– автоматизовані засоби моніторингу.

Крім того, архітектури SASE, які вирішують найпоширеніші сьогодні проблеми безпеки через додатки, що працюють за межами центрів обробки даних, можуть запропонувати додатковий захист.

7. Висновки

Граничні обчислення пов’язані з підтримкою туманних та хмарних рішень. Навіть за доступності компонентів для створення середовища, яке задовольняє більшості вимог, багато з них потребують тонкого налаштування або розширень API, щоб забезпечити більш оптимізоване та придатне цільове рішення. Вимоги до розгортання та тестування додатково визначаються для нових архітектурних моделей, і тому існуючі рішення необхідно покращувати, налаштовувати, а в деяких випадках розробити та впровадити з нуля. Основна проблема полягає в ефективному та ретельному аналізі нових концепцій і моделей архітектури, що розвиваються. Для такого аналізу необхідно ідентифікувати нові варіанти використання разом зі значеннями, які відповідають типовим обставинам застосування системи. Тестування забезпечує покращення архітектурних рішень і виявлення недоліків альтернативних рішень.

Існуючі моделі архітектури граничних обчислень охоплюють багато варіантів використання, однак вони потребують додаткових досліджень для деталізації необхідної функціональності розширення базових обмежень, визначення подальшого покращення рішень та документування передового досвіду.

Наприклад, рішення для Edge AI для промисловості передбачають розгортання глибоких нейронних мереж (DNN) зі спеціальними стратегіями для пом’якшення проблем, пов’язаних із великою кількістю пристроїв. Ці проблеми можуть бути вирішені з використанням активного навчання (англ. active learning, AL).

У аграрному секторі конвергенція традиційних технологій з AI забезпечує прийняття рішень у реальному часі, використання доказових аграрних практик, економію ресурсів та підтримку сталого розвитку.

У сфері охорони здоров'я інтелектуалізація граничного інтерфейсу підтримує спостереження за пацієнтами та забезпечення персонального медичного супроводу. Edge AI можна використовувати для аналізу даних пацієнтів, зібраних із різних джерел, включаючи електронні записи про стан здоров'я (англ. the electronic health record, EHR), портативні пристрої та медичні датчики для прогнозування ризику розвитку певних захворювань або медичних ускладнень.

Для побудови технологій розумного оточення граничні обчислення забезпечують інтелектуальне керування енергетикою, транспортом, комунальними послугами, а також інформування та реагування на надзвичайні ситуації.

На даному етапі розвитку інформаційних технологій різні об'єднання в IT-індустрії (як відкриті групи, так і напіввідкриті або закриті консорціуми, а також органи стандартизації) співпрацюють між собою при вирішенні проблем проєктування та тестування архітектури, щоб мати можливість задовольнити потреби різних випадків використання граничних обчислень.

Граничні обчислення на основі мікросхем ML забезпечують інтелектуальний аналіз даних безпосередньо на самих пристроях, що значно прискорює ефективність керування цільовими системами. Конвергенція AI та IoT підтримують автоматичне прийняття рішень, забезпечення безперебійної роботи складних систем керування та надання послуг, які відповідають потребам людей.

Конвергенція IoT та граничних мереж постійно поглибюється, однак існують деякі суттєві ризики для безпеки, пов'язані з мережами Edge та пристроями IoT, що робить безпеку Edge важливою технологією для забезпечення майбутнього IoT.

СПИСОК ДЖЕРЕЛ

1. Kovalenko O., Vishnevsky V., Kosolapov V. Towards Creating the Network of Situatonal Governance Centers and Decision Making Technologies in Distributed Environments. *2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*. Lviv-Slavske, Ukraine, 2020. P. 540–545. DOI: https://doi.org/10.1109/TCSET49122.2020.235491.

2. EdgeAI Consortium. URL: https://edge-ai-tech.eu/consortium/.

3. Iorga M., Goren N., Feldman L., Barton R., Martin M. Charif Mahmoudi NIST SP 500-325. Fog Computing Conceptual Model. 2018. 14 March. DOI: https://doi.org/10.6028/NIST.SP.500-325.

4. The Edge-to-Cloud Continuum. URL: https://www.computer.org/csdl/magazine/co/2020/11/09237349/1o8m4Fl5nUc.

5. The European Cloud, Edge and IoT Continuum Initiative. URL: https://eucloudedgeiot.eu/about/.

6. Khalyeyev D., Bureš T., Hnětynka P. Towards Characterization of Edge-Cloud Continuum. *Proc. of ECSA 2022 Tracks and Workshops*. 2023. P. 215–230. URL: https://doi.org/10.1007/978-3-031-36889-9_16.

7. Petersen A.M., Ahmed M.E., Pavlidis I. Grand challenges and emergent modes of convergence science. *Humanit. Soc. Sci. Commun*. 2021. Vol. 8. Article number 194. DOI: https://doi.org/10.1057/s41599-021-00869-9.

8. Roco M. Convergence-Divergence Process. *Handbook of Science and Technology Convergence* / W. Bainbridge, M. Roco (eds.). Cham: Springer, 2016. P. 79–93.

9. Kovalenko O. Knowledge Driven Cyber-Convergent Systems Based on Situational Agents. *2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT)*. 2022. P. 243–246. DOI: https://doi.org/10.1109/CSIT56902.2022.10000762.

10. Kovalenko O. Systems Convergence for Situational Control and Decision Making in Distributed Environments. *2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics,*

Telecommunications and Computer Engineering (TCSET). 2022. P. 344–347. DOI: https://doi.org/10.1109/TCSET55632.2022.9767006.
11. Athanasios Tziouvaras and Fotis Foukalas. Edge AI for Industry 4.0: An Internet of Things Approach. *Proc. of the 24th Pan-Hellenic Conference on Informatics (PCI '20). Association for Computing Machinery*. New York, NY, USA, 2021. P. 121–126. DOI: https://doi.org/10.1145/3437120.3437289.
12. Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, Yuan Gao A survey on federated learning, *Knowledge-Based Systems*. 2021. Vol. 216. P. 106775. DOI: https://doi.org/10.1016/j.knosys.2021.106775.
13. Edge AI Benefits, Use Cases & Deployment Models. URL: https://www.run.ai/guides/machine-learning-operations/edge-ai.
14. Introduction to Edge Computing in IIoT An Industrial Internet Consortium White Paper IIC:WHT:IN24:V1.0:PB:20180618 Edge Computing Task Group, ed. Stephen Mellor. URL: https://www.iiconsortium.org/pdf/Introduction_to_Edge_Computing_in_IIoT_2018-06-18.pdf.
15. The Essentials of SD-WAN Architecture: Advantages and Options. URL: https://www.sdxcentral.com/networking/sd-wan/definitions/software-defined-sdn-wan/essentials-sd-wan-architecture/.
16. What is Universal Customer Premises Equipment (uCPE)? URL: https://www.sdxcentral.com/networking/sd-wan/definitions/software-defined-sdn-wan/essentials-sd-wan-architecture/what-is-ucpe/.
17. vRan: How It Works & Why It Matters. URL: https://www.celona.io/network-architecture/vran.
18. Intel® Movidius™ Vision Processing Units (VPUs). URL: https://www.intel.com/ content/www/us/en/products/details/processors/movidius-vpu.html.
19. Accelerate AI development with Google Cloud TPUs. URL: https://cloud.google.com/tpu/?hl=uk.
20. NPU Ethos-U55. URL: https://www.arm.com/products/silicon-ip-cpu/ethos/ethos-u55.
21. Li W., Hacid H., Almazrouei E., Debbah M. A Review and a Taxonomy of Edge Machine Learning: Re-quirements, Paradigms, and Techniques. arXiv 2023, arXiv:2302.0857.
22. Edge AI Inference Platform. URL: https://www.aetina.com/products-features.php?t=337.
23. Is Secure Access Service Edge (SASE)? URL: https://www.zscaler.com/resources/security-terms-glossary/what-is-saseb.
24. How is Edge Security Helping Secure Devices that Use Edge AI? URL: https://xailient.com/blog/how-is-edge-security-helping-secure-devices-that-use-edge-ai/.

Стаття надійшла до редакції 12.08.2024