

Mesterséges intelligencia felhasználási esetek 5G hálózatokban

KOVÁCS BENEDEK¹, SZILÁGYI LÁSZLÓ²,
GERA ZOLTÁN, FÁBIÁN GÁBOR, CHARLES JOSE FERRARI³

¹Ericsson Magyarország Zrt., ²Ericsson, Inc., ³ELTE Informatikai Kar
benedek.kovacs@ericsson.com

Kulcsszavak: 5G-hálózatok, IoT, edge computing, elosztott felhő, neurális hálózatok, kiterjesztett valóság

Az 5. generációs hálózatok sokféle felhasználási eset számára fognak testre szabott megoldásokat adni, olyan területeken is, mint például az ipari automatizáció, okos mérés, a dolgok internete (IoT) és fejlett médiatechnológiát használó média. A cikk bemutatja, hogy milyen esetekben alkalmazható mesterséges intelligencia és gépi tanulás hálózatfelügyeletre és hálózatautomatizálásra, illetve bemutat egy pozíciót becslő, képfelismerő rendszert, mint alkalmazást 5. generációs hálózatra. Röviden ismerteti, hogy milyen módon támogatja az 5G-hálózat az úgynevezett edge computingot.

1. Bevezetés

A mobil telekommunikációs hálózatok már a kezdetek óta rendelkeztek intelligens megoldásokkal, eleinte leginkább automatizáláshoz köthető szabályokkal, később adaptív logikával. Összességében elmondható, hogy minden hálózati generáció a kor legmodernebb gépiintelligencia-technológiáját használja fel a megbízhatóság, robusztusság, biztonság és átviteli minőség biztosítására, automatizálására. Napjaink egyik legfelkapottabb témája a mesterséges intelligencia és annak különböző alkalmazási területei, melyet a neurális hálózatok fejlődése tett lehetővé.

A cikk célja, hogy bemutassa, hogy milyen helyzetekben releváns mesterséges intelligenciát alkalmazni 5. generációs mobil hálózatok esetén. A klasszikus gépi intelligencia felhasználási esetek mellett, melyek leginkább az intelligens hálózatok és automatizált felügyeleti rendszerekre fókuszálnak, bemutatunk lehetséges alkalmazási szintű megoldásokat, illetve ezek hálózati támogatását.

A második szakaszban leginkább a neurális hálózatok hőskoráról lesz szó, illetve olyan intelligens hálózati esetekről, amelyek a gépi tanulás felhasználásával valószínűsítették meg hálózati automatizációt. A következő szakasz a Big Data témakörét dolgozza fel és bemutatja, hogy milyen kihívásokat jelent a nagy mennyiségű (volume), sokféle (variety), gyors keletkezésű és rövid válaszidejű feldolgozást igénylő (velocity), sokféle minőségű (veracity) adat, és az, hogy milyen felhasználási esetekben segítenek a gépi tanulásos, neurális hálózatos megoldások. A negyedik rész a számítási kapacitási igényre koncentrálnak és egy példán keresztül bemutatja, hogy a nagy, felhőalapú adatközpontok, melyek tipikusak az 5. generációs hálózatok esetén, hogyan segítik a neurális hálózatok tanításával foglalkozó mérnökök munkáját. Egy alkalmazási szintű felhasználási esetet mutatunk be, amely képfelismeréssel segíthet intelligens közlekedési rendszerekben és egyéb intelligens meg-

oldásokban. Az utolsó szakasz bemutatja, hogy az 5G mobilhálózat a Distributed Cloud [1] technológiával hogyan támogatja az edge computingot és ezáltal a neurális hálózatokat használó alkalmazásokat.

2. Intelligens hálózatok és a mesterséges intelligencia fejlődése

Mesterséges intelligenciáról és intelligens hálózatokról különböző korokban és kontextusokban beszélhetünk. Az ötödik generációs hálózatokban az a különleges, hogy az elterjedőben lévő, neurális hálózatokon alapuló mesterséges intelligencia megoldásokat használja.

A mobil telekommunikációs hálózatok legelső verziói is tartalmaztak automatizálást, leginkább robusztusság és minőség biztosítására, például automatikus átkapcsolások meghibásodás esetén. Ezek a felhasználási esetek a felhasználó számára nem érzékelhetőek. Egyrészt az automatikus hálózatfelügyelet területét fedik le, automatizált újrakonfigurációval és beavatkozással (lásd [2]), másik részről pedig a hálózat elemeiben mélyen beágyazódva segítenek automatikus útvonalválasztással, adaptív túlterhelés-védelemmel. Intelligens hálózatoknak hívjuk továbbá azokat a mobilhálózatokat, melyek a felhasználónak nyújtanak intelligens hálózati szolgáltatásokat (ITU Intelligent Networks) [3]. Itt már megjelenik az 5. generációs hálózatokban bevett gyakorlat, hogy a hálózat bizonyos funkciói konfigurálhatóak harmadik fél számára.

A mesterséges intelligencia és gépi tanulás témaköréből a neurális hálózatok nyújtotta lehetőségeket fogjuk bemutatni, ezeken belül is a gépi tanuláson alapuló hálózatokat. Ezek kutatása és alkalmazása a múlt századig nyúlik vissza. A 21. század elején a neurális hálózatok elérhetővé váltak jórészt népszerű matematikai programcsomagok részeként, melyeket kutatók és mérnökök alkalmaztak, különböző témakörökben, például gyógyászati kutatásokban. Az elterjedést gátolta, hogy a fel-

használók számára elérhető személyi számítógépek nem voltak alkalmasak komplex hálózatok tanítására. Egy két-változós differenciálegyenlet-rendszer, a Lotka-Volterra egyenletek paramétereinek becslését egy kevesebb, mint 100 neuronból álló hálózattal lehetett elvégezni, melynek tanítása egy átlagos PC-n napokat vett igénybe [4].

Ezek a kis számú neuront tartalmazó hálózatok nem voltak alkalmasak komplex feladatok megoldására, például képfelismerésre. Az ilyen próbálkozások számára szuperszámítógépeket lehetett igénybe venni, ám ez drága volt és nem volt széles körben elérhető. Ehhez képest az emberi agyban lévő neuronok becsült száma százmilliárd, a legnagyobb publikált neurális hálózat 16 millió körüli neuronból áll, míg egy tipikus mai, képfelismerésre használható hálózat százezer neuront tartalmaz. Látható, hogy a tudomány fejlődésével és a számítási kapacitás növekedésével egyre komplexebb neurális hálózatokat építhetünk, azonban ezek a praktikus használatban még nagyságrendileg elmaradnak a biológiában természetesen előfordulóktól, ami további fejlődésüket vetíti elő.

A mesterséges neurális hálózatoknak több fajtája van, a telekommunikációs hálózatokban leginkább a mintalapú tanuló hálózatok különböző fajtáinak alkalmazása terjedt el. A jelenleg népszerű, *feature learning* hálózatok, a visszacsatolásos tanuláson alapuló hálózatok fejlesztett példái. A *feature learning* hálózat esetén a hálózat rejtett rétegei egy-egy értelmezhető funkciót végeznek el, így a mélytanulás (*deep learning*), azaz a sok rejtett réteg bevezetésével lehetővé válik a komponensenkénti tanulás vagy újratanulás. A tanítás időben is elterjedt, tehát tipikus, hogy egy általános problémára megtanított neurális hálózatot később, további minták segítségével egy adott problémára szabnak.

Összességében elmondható, hogy mára elérhetőek és egyre szélesebb körben elterjedtek olyan neurális hálózatok, melyek alkalmasak komplex problémák megtanulására, megértésére. Ezek jellemzően felhasználási esetek köré csoportosulnak és a *feature learning*, azaz előre tanított hálózatok esetén adott, konkrét problémák, például képfelismerés, hálózati hibaanalízis, biztonsági problémák felismerése, komplexitását képesek hatékonyan kezelni.

Az első körben bevezetett, modern neurális hálózatot alkalmazó felhasználási esetek többek között a lemorzsolódás előrejelzése és a hamis vagy nem biztonságos hálózati elemek kiszűrése voltak. A hálózati biztonság eléréséhez olyan neurális hálózatokat alkalmazunk, melyek kiszűrrik a különleges, vagy nem normális működési mintákat, automatikus beavatkozás vagy egy szakértő számára. A modern hálózatfelügyeleti rendszerekben alkalmazott neurális hálózatok figyelik, hogy milyen események vezethetnek hálózati forgalom vesztéséhez, kieséshez és így megtanulják, hogy milyen esetekben kell figyelmeztetést küldeni a felügyeletnek.

Ez a felhasználási eset azért érdekes, mert a jó és rossz esetek szétválasztása egyszerű szabályokkal leírható, így a tanulóhalmaz folyamatosan bővül. A *zero*

touch automatizáláson alapuló ötödik generációs hálózatok intelligenciáját is hasonló, fejlett neurális hálózatok segítségével tervezik majd elkészíteni [5].

3. Az adat, a Big Data és a hálózati analitika

A jelenleg elérhető tanuló rendszerek számára az egyik legnagyobb probléma az elegendő tanításra használható adat hiánya. Egy tipikus képfelismerési alkalmazás esetén, az előre tanított *feature learning* hálózatoknál is nagy mennyiségű, az adott felhasználási esetre specifikus tanítási adatra van szükség.

A modern telekommunikációs hálózatok felügyeletét és a felhasználói élmény biztosítását jelenleg a Big Data technológián alapuló támogató rendszerek végzik. A Big Data megelőzte a neurális hálózatok technológiáját és már a 4. generációs hálózatokra is jellemző, hogy az intelligens felügyeleti rendszereik a hálózat működése során mért teljesítményjellemzőket más adatbázisokkal összevetve képesek új értéket teremtő analitikai megoldásokra, például az Expert Analitika [6]. E rendszerek által felhasznált adatok például a felhasználókra jellemző anonim statisztikák, a terminálok típusai és gyártókra jellemző statisztikák, a hálózati elemek által generált teljesítményjelentések, naplók, a forgalmi statisztikák, az átvitel, hangminőség és a hálózati jelzésforgalom adatai.

Egy Voice over LTE (VoLTE) hívás esetén, az alkalmazásszintű, IP Multimedia Subsystem (IMS), Session Initiation Protocol (SIP) és Diameter protokollal jelzésforgalma hívásonként 200-300 üzenet (átlagos rendszer, átlagos felhasználás). Az üzeneteket térben és időben aszinkron módon figyelhetjük meg, mely nagy mértékben nehezíti a köztük rendszerszinten megfigyelendő korrelációt (a network timeoutok beállításától függően, egy azonnal kiszorgálandó üzenet akár több másodpercet is késleltethet). A megfigyelt üzenetekből levolt következtetéseket nehezíti, hogy szabványos rendszerek és alrendszereik nem tartalmaznak operáció szintű azonosítót (hiszen nem biztos, hogy egy adott operációhoz minden esetben minden alrendszer bevonása szükséges, egy-egy adatbázis lekérdezése például mehet aszinkron módon, cache használatával). Az egyes üzenetek korrelációja tehát korántsem egyértelmű probléma, és a gépi intelligencia alkalmazása így elkerülhetetlen.

Gépi tanulással felhasználási eset a jelzésforgalom elemzése, mely előnye azonban, hogy a tanító halmaz előállításához nem szükséges egyenként címkézni az elemeket, elég a hibakódokra vonatkozóan szabályokat létrehozni, ami sokszor igaz a felügyeleti alkalmazásokra. Ezzel a technikával a tanító halmaz minden esetben jó minőségű és folyamatosan bővül. Kifejezetten IMS-jelzésforgalom gépi tanulással elemzésével pár üzenet után megállapítható, hogy egy adott híváskezdeményezés milyen valószínűséggel lesz sikeres. Egy túlterhelt hálózat esetén, amikor döntenünk kell, hogy mely híváskezdeményezést (vagy csomagot) dobjuk el, kritikus, hogy a kiszorgált kérések sikeresek legyenek, hiszen a siker-

telen kiszolgálásnak értéke nincs. A végül sikertelen kiszolgálásra felhasznált hálózati kapacitást szaknyelven *blind load*-nak hívjuk.

A fent bemutatott felhasználási esetekhez kritikus a megfelelő hálózati kapacitás. A hálózatok virtualizálásának részeként a felügyeleti rendszereket is virtualizálják és a futtatási környezetük rugalmasan alakítható az aktuális igényekhez. A felhőalapú számítási és futtatási környezet több szempontból is kulcsfontosságú a neurális hálózatok tanításán alapuló megoldások számára. Egyfelől lényeges, hogy az adatot egy központi adattárházban tároljuk és érjük el, másfelől pedig fontos, hogy a tanítási folyamat elvégezhető olyan időszakokban, amikor olcsón áll rendelkezésre számítási kapacitás.

4. Az *Internet of Eyes* alkalmazás

A telekommunikációs alkalmazások és felügyeleti rendszerek tipikusan privát felhőben futnak, a nyilvános felhasználók számára viszont elérhetők az ún. *web-scale* cégek által nyújtott felhőszolgáltatások, például az Amazon Web Services, Microsoft Azure stb. A fejezet egy példaalkalmazást mutat be, demonstrálván, hogy milyen felhasználási eseteket tesz lehetővé az ötödik generációs mobilhálózat és hogyan lehet ezeket népszerű mesterségesintelligencia-eszközökkel megvalósítani. A példaalkalmazás egy demó keretében elkészült, munkacíme *Internet of Eyes*, és több tipikus kihívást is demonstrál, amelyek megoldása kulcsfontosságú az intelligens 5G hálózatokhoz: valós idejűség, elosztottság.

Az 5G-hálózatokkal kapcsolatban sok dolgot lehet biztosra venni már bevezetésük előtt is: nagy sávszélesség, kis késleltetés és sok eszköz együttes használata válik lehetővé segítségükkel. Ennél sokkal izgalmasabb kérdés azonban, hogy ha ezek a fejlesztések infrastruktúráisan is testet öltenek, milyen új szolgáltatásokra adnak majd lehetőséget és tulajdonképpen hogyan fogja az 5G egy átlagember életét befolyásolni. Az előző gondolat alapján érdemes főleg olyan szolgáltatásokat vizsgálni, melyek kis késleltetést, nagy sávszélességet és számításiigényt követelnek meg. Az *Internet of Eyes* elérhető, nagy sávszélességet igénylő eszközöket vesz célba: az IP-kamerákat. Több kamera képének egyidejű feldolgozásával, kis késéssel képes eseményeket, mozgást, objektumokat érzékelni, térben behatárolni, felismerni, majd az érzékelt információ gyors továbbításával beavatkozni. Példa lehet erre egy gépjármű megállítása, ha a kereszteződés felé olyan ütközőpályán érkezik egy másik közlekedő – mondjuk kerékpáros – mely a vezető szemszögéből nem látható, de a rendszer a kameráknak köszönhetően mind a kerékpárt, mind a gépjárművet és ezek mozgását érzékeli.

A rendszer fő összetevői

a) Adatforrás/Kamerák:

Nagy felbontású, folytonos videojelet továbbítanak. Rendszerint az ilyen kamerákon nincs meg a megfelelő teljesítmény komoly feldolgozás elvégzésére.

b) Párhuzamos és független feldolgozás:

Bizonyos műveleteket minden kamera képén lefuttatunk. Ide nemcsak zajszűrés és előfeldolgozás tartozik, de mozgásdetektálás, objektumkövetés és mélytanulással segített objektum-felismerés is.

c) Összegző feldolgozás/Pozíció-meghatározás:

A különböző kamerák jelfolyamából kinyert adatokat azonos koordinátarendszerbe illesztjük és meghatározzuk a résztvevő események/objektumok térbeli helyzetét, irányát.

d) Modellalapú reakció, ill. Beavatkozás:

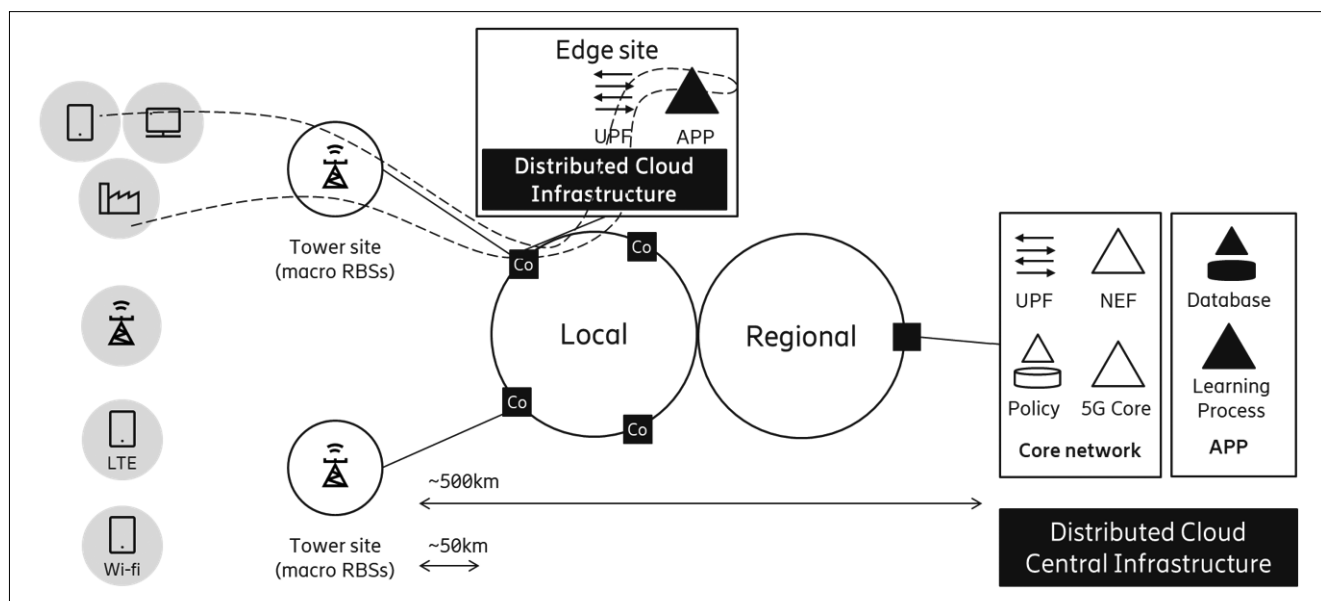
Itt történik meg a döntés, ami alapján beavatkozunk, vagy csinálunk valamit. Ez lehet aktív beavatkozás: egy autó leállítás az ütközés elkerülése érdekében, drón mozgásának megváltoztatása a lezuhanás elkerülése végett. Lehet azonban más rendszerek segítése is: egy kiterjesztett valóság kliens berajzolhatja az épületektől nem látszó tárgyakat a képre, így átláthatunk tárgyakon és valós időben láthatjuk, mi történik mögöttük.

e) 5G hálózat:

Lehetővé teszi, hogy a nagy sávszélességet igénylő kommunikáció ellenére a teljes visszacsatolási kör szinte valós időben történhessen meg. Továbbítja a jelet az adatforrásoktól a feldolgozó központok felé, ezek között, majd innen a cselekvés helyére, a beavatkozást végző eszközre. A teljes folyamat csak akkor ér valamit, ha a beavatkozás késedelem nélkül tud megtörténni. A videójel továbbítását RTP-protokollal és GStreamer [7] könyvtárral végezzük.

Videójel feldolgozására, zajcsökkentésre, objektumok követésére és mozgásérzékelésre OpenCV [8] programkönyvtárat használunk. A kis késleltetés és jó teljesítmény elérése érdekében a teljes rendszer C++-ban íródott. Ez alól a mélytanulással működő, objektum-felismerést végző TensorFlow [9] a kivétel. Ennek a komponensnek – és alapvetően minden jelenlegi mesterséges intelligencia megoldásnak – megvan az a hátulütője, hogy bár a legkomolyabb és legdrágább hardvert igényli, továbbra is a legnagyobb válaszidővel rendelkező egység a rendszerben. Ez annyira súlyos probléma a jelenlegi technikai színvonalon, hogy ha kizárólag ilyen algoritmusokra támaszkodnánk, akkor nem lenne elérhető valós idejű válaszidő a rendszer egészében a piacon kapható rendkívül drága, csúcscategóriás GPU-k használatával sem.

A közös koordinátarendszerbe helyezett rekombinációt egy saját fejlesztésű algoritmus végzi. A rendszer flexibilitását az adja, hogy valójában nem kell egy tárgyat sok kamerának látnia ahhoz, hogy tudjuk, hol van. Külön tudjuk kezelni a statikus kamerákat, melyek nem változtatják a helyzetüket. Ezek a rendszer indítása előtt kalibrálhatóak. A kalibráció lehetővé teszi, hogy a kamera saját helyzete alapján egyetlen képből is viszonylag jól lehessen becsülni, hogy a mozgás a tér mely részén, milyen térirányokban történik. Ilyen esetben egy másik kamera képe már csak pontosítja az adatokat. Dinamikus, azaz változó pozíciójú kamerák esetén pedig a többi kamera képe alapján ennek a kamerának a saját pozícióját is meg tudjuk határozni a látottak alapján.



1. ábra

Distributed Cloud, az applikáció által forgalmazott kritikus válaszidejű adatkomponens a hálózat szélén kerül feldolgozásra, az Edge Site-on: szaggatott adatút. Az Edge Site management és orkesztráció szempontból integrált része a felhő infrastruktúrájának, tehát tulajdonképpen egy elosztott felhőről, Distributed Cloud, beszélünk.

A konfiguráció az ábrán központban elhelyezkedő Network Exposure Function-on (NEF) keresztül végezhető el.

Ez történik egy kiterjesztett valóság kliensben (mondjuk mobiltelefon kliensen), amikor az ránéz a területre, majd ott kirajzolódnak a nem látszó objektumok is. Akárhogy mozgatjuk a mobilunkat, a berajzolt tárgyak együtt mozognak a látottakkal. A kiterjesztett valóság kliensünk C#-ban íródott, Unity3D [10] és Vuforia [11] keretrendszeret használ.

Fontos részlet a rendszer működésében a hálózati kialakítás is. Jellemzően mind az érzékelést végző kamera, mind a beavatkozást végző egység, illetve mobil kiterjesztett valóság kliens kis teljesítményű eszközök, melyeken a számítások nem lennének elvégezhetőek. Ugyanakkor az adatok nagy távolságra, a felhőbe való küldése túlságosan nagy késést okozna. Ezért van szükség a végesszközökhöz hálózati topológiában közel lévő, de nagy számítási teljesítménnyel rendelkező központokra ahhoz, hogy a kis késés ilyen jelentős terhelésnél is biztosított legyen. Az 5. generációs hálózatokban ezt a problémát a következő fejezetben bemutatott *edge computing* oldja meg, ahol a számítást végző egység nem a felhőből, hanem a szolgáltató topológiailag közel lévő egységéből választódik ki. Ezek a hálózat peremén rendelkezésre álló egységek azonban alkalmatlanak arra, hogy a komplex neurális hálózatok tanítását elvégezzék az adat és a kapacitás hiányában, így a tipikus munkafolyamat az, hogy a tanítást a központi felhőben, magát a valós idejű képfelismerést (inferenciát) pedig a hálózat peremén végezzük el.

5. Edge computing és az elosztott felhő

A 4. generációs hálózati architektúra jellemzője, hogy a klasszikus, HW alapú, vagy akár virtualizált hálózati eszközök egy központi adatfelhőben érhetőek el. Egy ti-

pikus közép-európai operátor országosan egy-két nagy adatközponttal rendelkezik. Az Amazon Web Services néhány nagy adatközpontot üzemeltet a világon, ezek közül többet Nyugat-Európában és egyet sem Kelet-Európában. Amennyiben a fent említett, alacsony késleltetésű alkalmazást szeretnénk elérhetővé tenni a mobilhálózatokban, úgy számolnunk kell azok késleltetésével.

A 1310 nm hullámhosszúságú fény egy tipikus Brand B (G.652) optikai kábelben történő terjedésének 489,34 $\mu\text{s}/100 \text{ km}$ sebességével számolva [12], a nyugat-európai adatközpont Magyarországról körülbelül 10 ms alatt érhető el, nem számítva a csatolók és útvonalválasztók által közbeiktatott, akár egyenként 1 ms késleltetést. Az országos adatközpontok tipikus átlagos elérése egy 4. generációs telekommunikációs hálózatban 10-50 ms. (Ezt az adatot lényegesen befolyásolja a rádiós és átviteli hálózat minősége és a felhasználás célja, például helyben telepített, privát hálózatok esetén nincs ez a típusú késleltetés.) Egyértelmű, hogy az egyes felhasználási esetek által megkövetelt 10-30 ms maximális késleltetés szükségessé teszi, hogy a csomagot feldolgozó alkalmazás a felhasználóhoz közel fusson. Ezt *edge computing*-nak nevezzük, mely technológiát lehetővé teszi az elosztott adatközpontok, melyek virtuális futtató környezetet kínálnak a hálózat peremén.

A 1. ábra bemutatja, hogyan illeszkedik egy tipikus edge computing alkalmazás egy elosztott felhő architektúrát támogató 5. generációs hálózatba. Az ábra bal oldalán található a terminál a rajta futó alkalmazás komponensekkel. Az 5G mobil hálózatokban bevezetett Next Generation Radio hálózati hozzáférés, konfigurációtól függetlenül az 1 ms nagyságrendű késleltetést is lehetővé teszi. A hálózat peremén futó elosztott felhő virtuális környezetének fizikai komponensei az ábrán *Central office*-ban (Co) helyezkednek el, ami tipikusan egy városi tele-

fonközpont modernizációjával alakítható ki, közel van a helyi antennához. Egy ilyen adatközpont tipikus mérete 5-25 pizzabox-szerver kapacitás. Könnyített hálózatfelügyelettel rendelkező infrastruktúrát adhat virtuális gép, konténer vagy funkció mint szolgáltatás típusú felhőalapú futtató környezetben hálózati eszközök és harmadik fél által fejlesztett alkalmazások vagy platformok számára. Azzal, hogy a virtuális infrastruktúra elérhető a hálózat szélén, nemcsak a hálózati késleltetést csökkentjük le a felhasználó készüléke és az alkalmazás között, hanem jelentősen csökkenthetjük az adatforgalmat, lehetőséget adhatunk tartalomszolgáltatások hálózatok lokális csatlósására.

Annak érdekében, hogy a felhasználók adatforgalma a megfelelő peremfelhőben futó applikációhoz kerüljön, a 3GPP hálózatok *local breakout* funkcióját az 5. generációs hálózati szabvány továbbfejleszti és az új funkcióval applikációk szerint lehet a forgalmat terelni és feldolgozni. Ez azt jelenti, hogy bizonyos applikációk forgalmát helyi feldolgozásra, másokét központi feldolgozásra csatolhatunk ki, de azt is, hogy egyes applikációk esetén más műveleteket végzünk a csomagokkal, amit a 3GPP szabvány *flexible mobile service steering*-ként definiál. Egy uplink videó esetén elképzelhető a *deep packet inspection* (DPI) szolgáltatás mellőzése, a kompresszió, vagy adott esetben konfigurálható, hogy titkosított vagy nem titkosított legyen a forgalom. A felhasználói adatot kezelő *User Plane Function* egyes elemei, melyek a 3GPP által definiált *flexible mobile service steering* komponensei, egy ilyen hálózatban akár szétosztott módon is futtathatók, tehát a hálózat különböző peremi vagy központi felhőjében lehetőséget adva intelligens teljesítmény optimalizálásra.

A 3GPP által szabványosított 5. generációs maghálózat (*5G Core*), architektúrájában jelentősen eltér a 4. generációtól (*Evolve Packet Core*). A tervezési szempontok közé tartozott a hálózati elemek, – ETSI terminológiában *Virtual Network Functions*, azaz virtuális hálózati funkciók – szolgáltatás központú architektúra (*service oriented architecture*) szerinti újratervezése, amit a szabvány *service based architecture*-ként ismer. A cél a programozhatóság és a rugalmasság támogatása volt. Az edge computing és elosztott felhő rendszerek szempontjából a *Network Exposure Function* kap kiemelt szerepet, mely lehetővé teszi megbízható harmadik fél által fejlesztett alkalmazások számára, hogy a hálózat bizonyos funkcióit konfigurálják. Ezen funkciók közé tartozik például az említett *application influence on traffic routing* programozható interfész, mely lehetőséget ad harmadik fél által készített alkalmazások számára a *local breakout* és *flexible mobile service steering* konfigurálására. Implementációs szempontból ez egy HTTP-alapú API, amelynek szolgáltatását a hálózat implementálja.

Az edge computing architektúrájú alkalmazások egyik központi kérdése, hogy mely modulok melyik fizikai helyszínen fussanak. Ennek eldöntésére egyszerű alkalmazások esetén kézi konfigurációt alkalmazhatunk, míg bonyolult alkalmazásokat intelligens optimalizáló rendszerek támogatnak majd [13].

6. Összefoglalás

Cikkünkben bemutattuk, hogy az 5G-hálózatokban milyen felhasználási esetek vannak a mesterséges intelligencia technológiák hálózaton belüli és applikációs rétegbeli alkalmazására.

Míg az első esetben, a hálózati automatizáció és az intelligens felügyeleti rendszerek természetes evolúciója következtében kezdődött meg a tanuló neurális hálózatok és egyéb mesterséges intelligencia megoldások használata, a másik, applikációs esetben a hálózat nyújtotta új lehetőségeket – *edge computing*, elosztott felhő, *local breakout*, *application influence on traffic routing* – fogják majd tudni kihasználni az 5G mobilhálózatra fejlesztett alkalmazások, aminek működését egy példával is demonstráltuk.

A szerzőkről



KOVÁCS BENEDEK vezető mérnök az Ericsson-nál. MSc-fokozatát mérnök informatikusként, PhD-fokozatát a Matematika és Számítástudományok Doktori Iskolában szerezte a Budapesti Műszaki és Gazdaságtudományi Egyetemen. Az Ericsson-ba 2005-ben szoftvertervezésként lépett be, majd rendszermérnökként részt vett több termék, többek között a Voice Over LTE szolgáltatás kidolgozásában. Specializációja a magas rendelkezésre állású rendszerek robusztussága és teljesítménye. Jelenleg aktívan dolgozik az 5G fejlesztésén, az ipari alkalmazások és az edge computing területén.



SZILÁGYI LÁSZLÓ a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai karán végzett, diplomamunkáját az ARM cambridge-i központjában végezte. Az Ericsson Magyarországnál szoftverfejlesztőként kezdett, rendszermérnöki és product owner-i szerepei után az elmúlt években innovációs projektek menedzselésével foglalkozott, IoT és edge technológiákra fókuszálva, az Ericsson Garage és az Ericsson technológiai szervezetének keretein belül. Jelenleg az Ericsson kaliforniai Santa Clara-i irodájában dolgozik, az új üzleti területek értékesítési részlegén az edge-technológiáért felel.



GERA ZOLTÁN az ELTE Informatikai Karán tanársegéd, valamint kutatás-fejlesztési projektek vezetője. Két jelentős magyar startup (NNG, Prezi) első négy évének részese volt fejlesztőként, csapatvezetőként. A kutatás és innováció motiválja, USA-ban bejegyzett szabvánnyal is rendelkezik. Szívéügye az informatikai szakma jobbá tétele, ezért igyekszik alma mater-én minél többet visszadni a közösségnek.



FÁBIÁN GÁBOR az ELTE IK-n tanársegéd, az intézmény Doktori Iskolájának doktorjelöltje. Több projekt megvalósításában végzett kutató-fejlesztő munkát, melyek az informatika olyan területeit érintették, mint a hangfeldolgozás, képfeldolgozás, vagy a 3D rekonstrukció. Kutatásában a számítógépes grafika matematikai módszereivel foglalkozik, de szívesen fordítja figyelmét más területek felé is. Gyakorta tart tudománynépszerűsítő előadásokat, ugyanúgy fontosnak tartja a tudományos módszerek és eredmények magas szintű oktatását, mint azok vonzóbbá tételét.



CHARLES JOSE FERRARI három éve PhD-hallgató az ELTE IK-n, melynek keretein belül az 5G hálózatok szoftverfejlesztési applikációit kutatja, kifejezetten az edge computing applikációk és környezet területén. Tíz éves szoftverfejlesztői tapasztalattal rendelkezik, többek között saját startup cége keretein belül végzett szakmai munkát. Számítógépes Tudományok MSc-oklevelét Innováció és Vállalkozás kiegészítéssel az Európai Innovációs és Technológiai Intézet duális képzésében az ELTE-n szerezte.

Hivatkozások

- [1] C. Boberg, M. Svensson, B. Kovács, "Distributed cloud – a key enabler of automotive and industry 4.0 use cases", November 2018, <https://www.ericsson.com/en/ericsson-technology-review/archive/2018/distributed-cloud>
- [2] M. Svensson, M. Agarwal, S. Terrill, J. Wallin, "Open, Intelligent and Model-driven: Evolving OSS", 7 Februar 2019, <https://www.ericsson.com/en/ericsson-technology-review/archive/2018/open-intelligent-and-model-driven-evolving-oss>
- [3] ITU Intelligent Networks, <https://www.itu.int/rec/T-REC-Q.1200/en/>
- [4] B. Kovacs, J. Toth, "Estimating Reaction Rate Constants with Neural Networks", 2007.
- [5] E. Fersman, J. Forgeat, R. Cöster, S. K. Mohalik, V. Berggren, "Artificial Intelligence and Machine Learning in Next-generation Systems", Ericsson White Paper, June 2018, <https://www.ericsson.com/en/white-papers/machine-intelligence>
- [6] Expert Analytics, <https://www.ericsson.com/en/portfolio/digital-services/automated-network-operations/analytics-and-assurance/expert-analytics>
- [7] GStreamer média-streaming könyvtár, <https://gstreamer.freedesktop.org/>
- [8] OpenCV számítógépes látás könyvtár, <https://opencv.org/>
- [9] TensorFlow gépi tanulás keretrendszer, <https://www.tensorflow.org/>
- [10] Unity3D játékmotor, <https://unity3d.com/>
- [11] Vuforia kiterjesztett valóság keretrendszer, <https://www.vuforia.com/>
- [12] K. Miller, Calculating Optical Fiber Latency, 9 Januar 2012, M2 Optics Inc., <http://www.m2optics.com/blog/bid/70587/Calculating-Optical-Fiber-Latency>
- [13] A. Reale, Kiss P., C. Ferrari, Kovacs B., Szilagy L., Toth M., "Application Functions Placement Optimization in a Mobile Distributed Cloud Environment", In: Studia Informatica, no.2, pp.37–52, 2018, http://www.studia.ubbcluj.ro/arhiva/cuprins_en.php?id_editie=1155&serie=INFORMATICA&nr=2&an=2018



www.hte.hu

HTE
INFOKOM
MEDIANET

KLUBÉLET
JOURNAL
HÍRADÁSTECHNIKA
KONFERENCIÁK SZERVEZÉSE
PROJEKTMENEDZSMENT FÓRUM
SZAKMAI DÍJAK
ODAITÉLÉSE

FOLYÓIRATOK

IEEE
ÉS MÁS
TÁRSZERVEZETEK

TEVÉKENYSÉG
TÁMOGATÁSA

INFOCOMMUNICATION
KIEGYENSÚLYOZOTT
SZAKMAPOLITIKAI,
SZAKMAI
VÉLEMÉNYALKOTÁS
NEMZETKÖZI
KAPCSOLATOK

info@hte.hu

HÍRKÖZLÉSI ÉS INFORMATIKAI
TUDOMÁNYOS EGYESÜLET (HTE)
HTENET INNOVÁCIÓS NONPROFIT KFT

1051 Budapest, Bajcsy-Zsilinszky út 12.
Tel.: 353 1027
E-máll: info@hte.hu