híradástechnika

1945 VOLUME LXXVII. 2022

hírközlés ▪ informatika

1

HTE Infokom 2021

A Hírközlési és Informatikai Tudományos Egyesület folyóirata

Tartalom / Contents

Hírközlési és Informatikai Tudományos Egyesület
www.hte.hu
Elnök: Vágujhelyi Ferenc
H-1051 Budapest, Bajcsy-Zsilinszky út 12., 5. em./502.
Tel.: 353-1027 • e-mail: info@hte.hu

Főszerkesztő
SZABÓ CSABA ATTILA (BME, Hálózati Rendszerek és Szolgáltatások Tanszék)

Felelős kiadó: NAGY PÉTER
HU ISSN 0018-2028
Layout: MATT DTP Bt. • Nyomda: FOM Media

A folyóirat támogatói

Nemzeti Együttműködési Alap
Miniszterelnökség
Bethlen Gábor Alapkezelő Zrt.
Emberi Erőforrások
Petőfi Kulturális Ügynökség

www.hiradastechnika.hu

HTE Infokom 2021

2021. november 3-4-én huszonharmadik alkalommal került megrendezésre a Hírközlési és Informatikai Tudományos Egyesület szervezésében az Infokommunikációs Hálózatok és Alkalmazások Konferencia és Kiállítás, a *HTE Infokom*. A helyszín a Danubius Hotel Helia volt.

A rendezvény a hazai infokommunikációs szakma kimagasló eseménye, elismert tudományos-szakmai fóruma. Résztvevői és előadói elsősorban az IKT területén tevékenykedő cégek vezető szakemberei, kutató-fejlesztő projektvezetői, műszaki-technológiai döntéshozói. A konferencia célja, hogy lehetőséget teremtsen az infokommunikációs piac változásainak megismerésére, a legújabb műszaki megoldások, hálózat-, szolgáltatás- és alkalmazásfejlesztési elképzelések közzétételére, tapasztalatok kicserélésére, az együttműködés elmélyítésére, a személyes és közvetlen kapcsolatok kialakítására.

Számunk cikkeit az Infokom 2021 előadásaiból válogattuk össze. A cikkek sorrendje követi a konferencia szekcióinak sorrendjét.

Kákonyi István (Cisco) „A távközlési infrastruktúra fejlődése a digitális átállás és a felhőalapú technológiák korában” címmel bemutatja a szolgáltatói hálózatok fejlődésének legfontosabb szempontjait. A távközlési szolgáltatóknak meg kell küzdeniük az egyre növekvő sávszélesség-felhasználással és a felhőalapú alkalmazások elterjedésével. Változnak a forgalmi minták is, a hálózatfejlesztésnek pedig a szűkülő CAPEX- és OPEX-feltételek ellenére is meg kell történnie. Egyszerűsítésre és automatizálásra van szükség a hálózatban.

Czintula György (Pro-M) „Készenléti közszolgáltatás biztosítása járványhelyzetben” írása azt ismerteti, milyen járvány okozta kihívásokkal kellett szembenéznie a kormányzati célú hírközlési szolgáltatónak az EDR-közszolgáltatás fenntartásában. A COVID-19 járvány a felhasználóknál megnövekedett többletfeladatokat eredményezett, amelyek az EDR-szolgáltató részéről gyors válaszokat tettek szükségessé. A cikk vázolja az üzletmenet-folytonosság érdekében meghozott szervezeti, rezsim- és adminisztratív intézkedéseket. Összefoglalót ad a járványkezelés eredményességéről, kitekintve a COVID-típusú kihívások jövőbeli kezelésére.

Az edge computing napjaink egyik legnépszerűbb technológiája. Bár a koncepció évek óta létezik, az 5G- és IoT-rendszerek megjelenésével került ismét előtérbe. *Kovács Benedek* (Ericsson) „Az edge computing, mint disztruptív technológia” című írása bemutatja, hogy miért tekinthetjük disztruptív technológiának. Áttekinti az edge computing főbb felhasználási eseteit, az iparági szereplőket, motivációikat és a lehetséges értékláncokat. A cikk példái mind üzleti, mind technológiai szempontból bemutatják az edge computingot.

Bartolits István (NMHH) „Mit ad az átlagfogyasztónak az 5G?” írása arra keresi a választ, hogy a vertikumok számára sok előnyt ígérő 5G-hálózatok mit nyújtanak az átlagos fogyasztók számára most és a jövőben. Bemutatja az 5G-rendszerek jelenlegi piaci helyzetét, majd kitér a piaci bevezetés nehézségére. A már az „5G non standalone” rendszer alatt igénybe vehető alkalmazások mellett rámutat a távolabbi jövő, az „5G standalone” rendszer sokoldalú szolgáltatásaira.

A „Felhők biztonsági kérdéseinek aktualitásai” cikk szerzője, *Kovács Zoltán* (Vodafone) rávilágít a felhőalapú rendszerek megkerülhetetlenségére, példákat hoz a speciálisan felhőalapú rendszereknél jelentkező biztonsági problémákra, majd bemutat egy lehetséges módszertant annak eldöntésére, hogy egy felhőalapú rendszer teljesíti-e a minimálisan elvárt biztonsági szintet. Példákat hoz a mesterséges intelligencia és annak felhőben működő megoldásainak védelmi célú használatára, és bemutatja annak támadó oldalon történő felhasználását is.

A távközlési iparágban használt infrastruktúra- és platform-megoldások egyre inkább konvergálnak az egyéb IT-megoldások által használt platformok irányába. Ez rengeteg előnnyel jár technikai és üzleti szempontból is, ugyanakkor ezzel új fenyegetések jelennek meg a távközlés területén. *Csordás Gábor* (Nokia) „Cloud-infrastruktúra és a rajta futó telekom-applikációk aktuális biztonsági kihívásai” címmel ezeket tárgyalja, sorra véve a privát és publikus cloud-rendszerek általános sebezhetőségét, a nyílt forráskódú szoftverek beépítését az infrastruktúrába és a telekom-applikációkba, illetve a privát cloudból a publikus cloud-rendszerek felé elmozdulás miatti változásokat.

Nelson Francisco–Bordás Csaba (MediaKind) „A mesterséges intelligencia és az új kódolási eljárások szerepe a videófeldolgozás fejlődésében” című cikkükben bemutatják, hogy ha a videotömörítő rendszerek tervezésében kihasználják a gépi tanulás lehetőségeit, akkor segítségükkel olyan valós idejű mesterséges intelligencia-vezérelt kódolási döntéseket lehet hozni, amelyek sokkal hatékonyabbnak bizonyulnak, mint bármely ember által meghatározott heurisztika vagy algoritmus.

Az IEEE Hungary Section támogatásával, illetve az Ericsson és a Nokia ipari partnerségében az Infokom 2021 konferencián először került megrendezésre a tudományos szekció, *Farkas Károly* (Netvisor, BME) szervezésében. Ennek keretében mutatkoztak be a *HTE Diplomaterv és Szakdolgozat Pályázat* kategórianyertesei, akiknek összefoglalóiból almanach készült, amelyet a jelen különszámba is beszerkesztettünk. A munkákat Farkas Károly előszava mutatja be.

Szabó Csaba Attila
főszerkesztő

A távközlési infrastruktúra fejlődése a digitális átállás és a felhőalapú technológiák korában

KÁKONYI ISTVÁN
Cisco Systems Magyarország Kft.
ikakonyi@cisco.com

Kulcsszavak: SDN, „merchant silicon", automatizálás, virtualizáció, SDN-vezérlő, adatsík, IPoDWDM, konténeralapú virtualizáció

A távközlési szolgáltatóknak meg kell küzdeniük az egyre növekvő sávszélesség-felhasználással (amit a Wi-Fi, 5G és más szélessávú technológiák indukálnak) és a felhőalapú alkalmazások elterjedésével. Nemcsak a sávszélesség, hanem a forgalmi minták is változnak. Másrészt a hálózatfejlesztésnek a szűkülő CAPEX- és OPEX-feltételek ellenére is meg kell történnie. Egyszerűsítésre és automatizálásra van szükség a hálózatban.
A cikk bemutatja a szolgáltatói hálózatok fejlődésének legfontosabb szempontjait.

1. Bevezetés

A távközlési szolgáltatók újabban komoly kihívásokkal néznek szembe. A felhasználók egyre nagyobb sávszélességű technológiákat szeretnének használni (ez igaz a vezetékes és a mobil hálózatokra is), emellett terjednek a felhőalapú szolgáltatások. Ezek a szolgáltatások átrajzolják az eddig megszokott forgalmi irányokat, mert a hagyományos „észak-dél" (letöltés) viszonylatok helyett a „kelet-nyugat" (adatközpontok közötti és az adatközpontokon belüli) irányok is nagyon fontosak lesznek. A fenti követelményeknek megfelelő hálózatot pedig változatlan vagy egyenesen csökkenő költségekkel kell megvalósítani.

Cikkünkben áttekintjük az ilyen hálózatok építőelemeit és architektúráját.

Alapvetően három kérdést fogunk vizsgálni:

- A hardverelemek fejlődését, illetve azt, hogy milyen lehetőség van egyszerűbb, olcsóbb hálózati elemeket létrehozni.
- A szoftvertechnológia, illetve a virtualizáció és az automatizálás fejlődését. Sok előrelépés történt az eszközök menedzsment-interfészeinek hatékonyabbá tétele terén (orchestration, Netconf-protocol, Yang-adatmodellek). Sok funkciót meg lehet teljesen szoftveres alapon valósítani és virtualizált környezetben használni.
- Az architektúra fejlődését, illetve egyszerűsödését, és ezt egy létező, feltörekvő technológián keresztül (Segment Routing) mutatjuk be.

2. A router/switch hardver fejlődése

A nagy teljesítményű routerek, amelyek szolgáltatói környezetben is alkalmazhatóak, hosszú ideig azonos séma szerint készültek: a gyártó néhány év alatt kifejlesztett egy ASIC-generációt, amely képes volt a több millió IP-prefix, a lapkánként több Tbit/s sávszélesség kezelésére, és e köré az ASIC köré épült fel a hardver. Az operációs rendszer általában valamilyen Linux disztribúcióból származott, és a HAL (hardware abstraction layer) segítségével több platformra is adaptálható volt. Egy ilyen folyamat költséges és hosszú, az eredménye azonban olyan (drága...) eszköz lehet, ami bizonyos tulajdonságaival felülmúlja a konkurens termékeket.

A piaci igényt felismerve bizonyos gyártók elkezdtek általános használatra alkalmas, de nagy skálázhatóságú chipeket fejleszteni. Ezek először az adatközponti eszközökben terjedtek el, ma már szolgáltatói routerekben is megtaláljuk őket. Ilyeneket fejleszt és gyárt pl. a Broadcomm, a Marvell vagy az Innovium. Az eredmény egy polcról levehető architektúra („merchant silicon"). A gyártó általában egy API-t is ad a chipekhez, így az alkalmazók könnyen adaptálhatják a már meglevő szoftvereiket. Az eredmény: rövidebb fejlesztési idő, olcsóbb eszközök, alacsonyabb energiafelhasználás. A dolognak természetesen árnyoldala is van: a „merchant silicon" routerek hasonló paraméterekkel fognak rendelkezni, a gyártók csak a szoftver segítségével tudják megkülönböztetni a termékeiket.

A fent leírt építőelemek és „merchant silicon" routerarchitektúrák fejlődése töretlen, ma már minden távközlési szolgáltatói igényt ki tudnak elégíteni (aggregáció, gerinchálózat, ISP peering).

Az *1. ábra* összefoglalja az egyik népszerű „merchant silicon" ASIC különböző változatainak a paramétereit.

Meg kell még jegyezni, hogy a 100 és 400 GE interfészek elterjedésével az optikai interfészek, modulok ára egyre magasabb hányadot képvisel a router teljes árában. A koherens WDM-technológiák nagyon sokat fejlődtek, ma már 400 Gbit/s sebességű optikai modulok is elérhetők, koherens transceiverrel.

Vannak olyan gyártók, amelyek kizárólag a „merchant silicon"-csipekre alapozzák a termékeiket. Vannak olyanok is, amelyek fejlesztik a saját ASIC generációikat, és emellett sokkal olcsóbban kínálnak „merchant

silicon”-alapú termékeket is. A folyamat kiélezte a gyártók közötti versenyt, a vásárlók számára pedig előnyösebb pozíciókat eredményezett.

3. Szoftver, virtualizáció, automatizálás

A routerek szoftverarchitektúrája szintén drámaian fejlődött az elmúlt években. A szabványosításra való törekvés mellett ehhez jelentősen hozzájárult az Open Source technológiák fejlődése és adaptálása.

Az eszközök menedzselhetőségében igazi áttörés következett be: az elavult, nehézkes, vállalati használatra kifejlesztett SNMP-protokoll helyett jött a Netconf, amely tranzakcióalapú, és képes absztrakt adatmodellekkel dolgozni (YANG). Az eredmény az, hogy az egyes eszközök, sőt komplett szolgáltatások is leírhatók YANG-modellekkel. Ez lehetővé teszi különböző gyártók termékeinek ugyanabban a hálózatban való alkalmazását. Ha a megfelelő YANG-modellek rendelkezésre állnak, akkor a szerviz-logikát tartalmazó „orchestrator” az összes eszköz számára el tudja küldeni a szükséges Netconf-parancsokat és ellenőrizheti is azok sikeres végrehajtását. A Netconf-protokoll mellett más, egyszerűbb API-k is elterjedtek a routerek világában, pl. RESTCONF vagy a JSON-alapúak is.

A hálózat működési paramétereinek vizsgálata és azok elemzése is új szintre került. A hagyományos SNMP-alapú lekérdezések helyett ma már elterjedt a „streaming telemetry”, amely gyakorlatilag azt jelenti, hogy a felügyeleti rendszer meghatározhatja, hogy milyen paraméterekre kíváncsi és ezt az eszközök folyamatosan küldik. Ezt egyes esetekben a hardware is támogatja. Így akár IP-flow-szintű adatok is kinyerhetők. A legelterjedtebb protokollok a már említett Netconf és a gRPC (Google RPC).

A szolgáltatásmenedzsment csúcsán egy olyan szoftver van, amelyet orkesztrátornak hívunk. Ez a szoftver képes megérteni a szolgáltatási vagy alkalmazási logikát, és képes ezeket az egyes eszközöket – mindegy, hogy hardver vagy virtualizált elemekről beszélünk – úgy konfigurálni, hogy a kívánt szolgáltatás létrejöjjön.

Sok olyan hálózati funkció van, amit szoftveralapon is meg lehet valósítani. Az ezt leíró szabványokat, architektúrákat egységesen Network Function Virtualizationnak (NfV) hívja a szakirodalom.

Fontos kérdés, hogy mit érdemes virtualizálni? Természetesen minden kontrollsík-funkciót (routing protokollok, traffic engineering stb.) lehetséges, általában olyan feladatokat érdemes, ahol sok és komplex számítási igény van. Az SDN-technológia hőskorában (úgy 10 éve) úgy gondolták, hogy a teljes kontrollsík virtualizálva lesz és a routert és a kontrollert össze kell kapcsolni egy protokollal, ami csak a csomagtovábbításra vonatkozó információt továbbítja. Ez volt az OpenFlow. Korlátozottan terjedt el, mert kiderült, hogy jobb, ha megmarad a router kontrollsíkja, de azt szabványos interfészekkel ki kell nyitni, és így előtte elképzelhetetlen szolgáltatásokat lehet bevezetni (erről a későbbiekben lesz még szó). A szolgáltatói router, különösen a gerinchálózati eszközök az infrastruktúra kritikus elemei. Célszerű, ha meghagyjuk őket valamennyire autonóm üzemmódban. Ezt a koncepciót egyesek „hybrid SDN”-nek hívják.

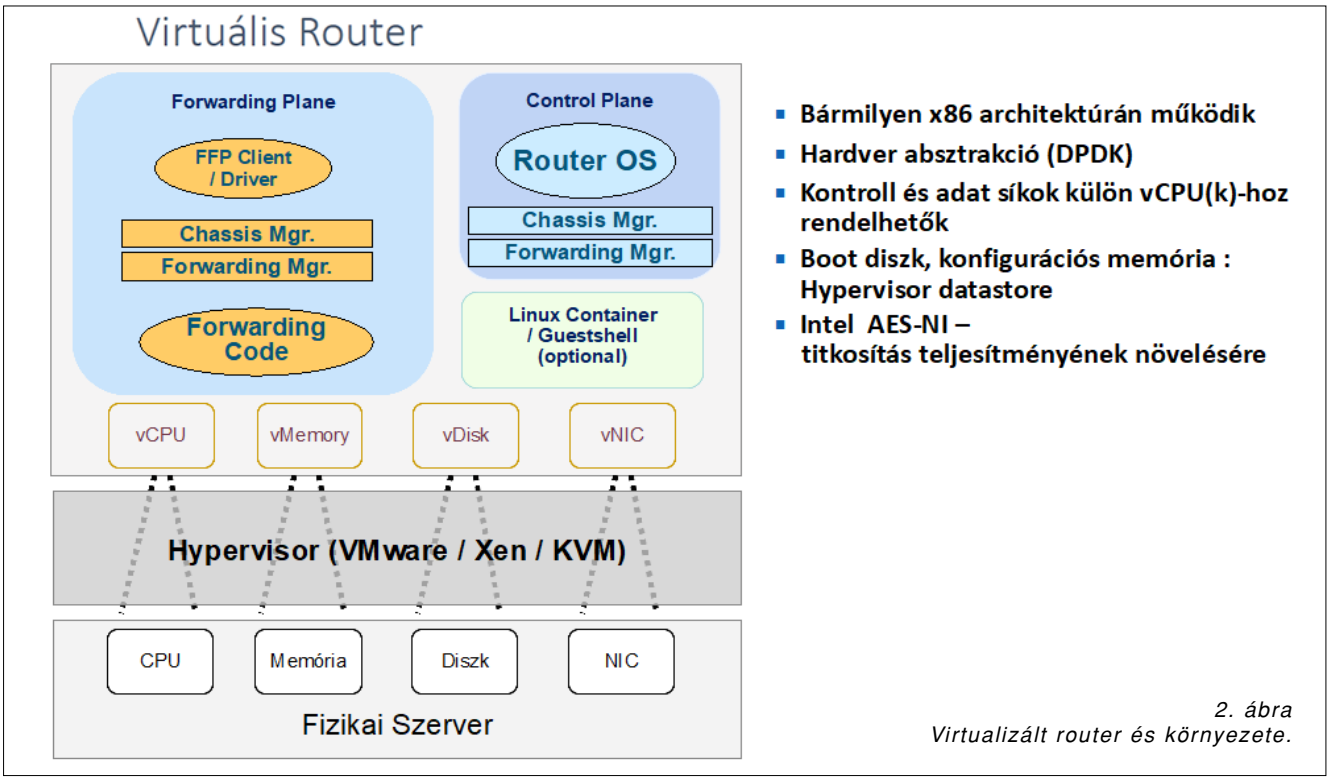
Az adatsík nehezebben virtualizálható, bár ma már ez is lehetséges. Az Intel DPDK (Data Plane Development Kit) megjelenése áttörést jelentett: az általános CPU-k alkalmassá váltak IP-csomagok hatékony továbbítására is. A fejlődés valóban robbanásszerű: ma már minden komolyabb gyártónak van szoftveralapú routere, sőt ma már a második generációról beszélhetünk. Az első generációs eszközök általában egy meglevő routercsaládból indultak ki, és az abban levő hardvert emulálták egy szoftverréteg segítségével, ami egy szabványos CPU-n futott. A második generáció már a szabványos szerverarchitektúrák (Intel, AMD) sajátosságait figyelembe véve sokkal nagyobb teljesítményre képes. Ma nem ritka a CPU-socketenkénti 10 Gbit/s teljesítmény sem.

A *2. ábrán* egy szoftveralapú, teljesen virtualizált router architektúráját láthatjuk.

1. ábra

Egy népszerű „merchant silicon” ASIC különböző típusainak jellemző paraméterei (Broadcom BCM88 sorozat).

	Jericho	Jericho +	Jericho2	Jericho2C	Jericho2C+
Sávszélesség Gbps	720	900	4,800	2,400	7,200
Teljesítmény PPS	720M	835M	2B	1B	2.83B
Interfészek	24x 25G+36x 12.5G	48x25G+24x12.5G	96x 50G	32x50G+96x25G	144x 50G
Fabric Interfész	36x 25G	48x 50G	112x 50G	48x 50G	192x 50G
Chip technológia	28nm	28nm	16nm	16nm	7nm
OCB	16MB	16MB	32MB	32MB	32MB
Buffer	4GB (GDDR)	4GB (GDDR)	8GB (HBM)	4GB (HBM)	8GB (HBM)
Virtual output queue	96K	96K	64K per core	128K per core	256K per core
Számlálók	256K	256K	192K	384K	384K
Fogyasztás	120W	150W	300W-350W	150-200W	450W
PTP támogatás (on Chip)	Class B	Class B	Class B	Class C	Class C
Macsec	No	No	No	No	Yes



2. ábra
Virtualizált router és környezete.

Az NfV általában valamilyen jól bevált virtualizációs környezetet használ (VMware, Openstack és hasonlók). A virtualizációs rendszer és az erőforrásmenedzsment gondoskodik arról, hogy mindig megfelelő számú virtuális router, load balancer, firewall stb. álljon rendelkezésre az adott terhelésnek megfelelően. Az NfV előnye a hatalmas rugalmasság: ha bővíteni kell a rendszert, ugyanolyan szabványos szervereket kell venni és üzembe helyezni. A virtualizáció legújabb iránya a konténeralapú virtualizáció (az egyik legelterjedtebb konténertechnológia a Docker). A konténeres technológia előnye, hogy a Hypervisor által okozott többlet-CPU- és memóriaigény nagyrészt kiküszöbölhető. A konténeres rendszerek erőforrás-menedzsmentje, policy-menedzsmentje és hálózati integrációja ma már megoldott.

Az NfV egyik sikertörténete a mobil internet egyik alapvető építőeleme, a Mobile Packet Core. Erre alkalmazhatók a fentebb leírtak: CPU-intenzív, sok előfizető forgalmát kell egyszerre feldolgozni, és a transzporthálózat bevezeti a forgalmat az adatközpontba, ahol azt fel lehet dolgozni. A MPC-megoldások eleinte még célhardvereket használtak, aztán a fent leírt fejlesztések már lehetővé tették a tisztán virtualizált („cloud native”) megvalósítást is.

Ma már az 5G-mobilhálózatok úgynevezett „cloud native” architektúrát alkalmaznak. Ez azt jelenti, hogy a rádiós hardvert kivéve minden funkció virtualizálva van. Ez hatalmas előnyt jelent a fejlesztőnek és a vásárlónak/üzemeltetőnek is. Azonos hardverelemekből kell építkezni, igen jól skálázható a megoldás és nagyon magas üzembiztonság érhető el.

A programmatikus interfészek, a streaming-telemetria és az SDN-kontroller alkalmazása lehetővé teszi bizonyos funkciók automatizálását, csökkentve ezzel a hálózat üzemeltetésének költségeit. Nagyon sok alkalmazási példa van erre, ezek közül egyet emelnénk ki. *Az IP és az optikai hálózat kontrollsíkjának integrálása* eddig nem látott funkciók megvalósítását teszi lehetővé. Ehhez a következő technológiák szükségesek:

- Koherens DWDM-átvitel, hangolható transceiverekkel.
- IP- és DWDM-integráció a routereken.
- Programozható hullámhosszkapcsolók a DWDM-hálózatban.
- Programmatikus interfészek az IP- és DWDM-doménben.
- SDN-kontroller.

Ha a fentiek rendelkezésre állnak, az SDN-kontroller képes észlelni a routerből érkező telemetria-jelekből, hogy egy adott DWDM-linken rosszabbodik az átvitel minősége (romló SNR, romló hibaarány). Ha ezek az értékek a beállított tartományokból kiesnek, a rendszer képes arra, hogy új optikai adatutat keressen (hisz az optikai szálak topológiája is adott), és ezt az optikai doménben automatikusan kialakítsa. Ezek után az IP-forgalom más irányban, más topológián keresztül továbbítódik. A mai korszerű rendszerekben az átkapcsolás n x 10 sec nagyságrendű lehet.

A fenti művelet az SDN-technológia alkalmazása nélkül kézi beavatkozást és akár napokat is igényelhet.

4. Korszerű szolgáltatói hálózati architektúra

A távközlési szolgáltatók ma alapvetően MPLS-technológiát használnak a hálózataikban. Az MPLS egy absztrakciót használ, az egyes irányokat (IP-prefix), szolgálta-

tásokat vagy tunneleket egy-egy címkével (label) azonosítja. A hozzárendelést az LDP-, az RSVP-, vagy a BGP-protokoll végezheti. Az eredmény egy stabil, skálázható, de komplikált hálózat. Az idők folyamán igény mutatkozott az egyes forgalomtípusok megkülönböztetésére és egy adott útvonalon történő továbbítására. Ezt a problémát képes megoldani a Traffic Engineering, de ez egy új elem bevezetésével járt a protocol stack-be: ez az RSVP (Resource Reservation Protocol). Az MPLS Traffic Engineering nem tudott széles körben elterjedni, mert az újabb elem a protocol stack-ban és a routereken jelentkező plusz CPU- és memóriaigény nehézkessé tette a használatát (az RSVP-protokoll „stateful”: az összes részt vevő routeren nyilván kell tartani a tunneleket).

A Segment Routing (továbbiakban SR) [1] az MPLS adatsíkjának változatlanul hagyásával (ezáltal natív IPv6-hálózaton is működik, ennélfogva jövőálló!), de a kontrollsík jelentős egyszerűsítésével jött létre. Ma már gyakorlatilag minden vezető gyártó támogatja, ezért kvázi szabványként is tekinthetünk rá. A Segment Routing legfontosabb tulajdonságai a következők:

- MPLS- vagy IPv6-adatsík (ez utóbbi esetén nem label stack, hanem IPv6-címek rendezett listája van az IPv6 routing-header-ben).
- Source routing, a label impozíciót végző router által az IP-csomagra csatolt „label stack” alapján történik a forgalom irányítása.
- Nincs LDP-protokoll, a címkék (Segment ID) allokálása és terítése az IGP-protokoll feladata. IS-IS és OSPF kiterjesztések segítségével történik.
- Minden szolgáltatás, ami MPLS-hálózaton működik, SR-hálózaton is fog (L2VPN, L3VPN) működni, minden változtatás nélkül.
- Nem szükséges újabb protokoll a Traffic Engineering megvalósításához. SDN-kontrolleren keresztül, szabványos API-n (Path Computation Element Protocol) bármilyen „tunnel” beprogramozható.

A *3. ábra* mutatja az MPLS és a SR közötti különbségeket. Ahhoz, hogy az SR-technológia működését pontosan megértsük, talán érdemes röviden áttekinteni a két mechanizmus működését.

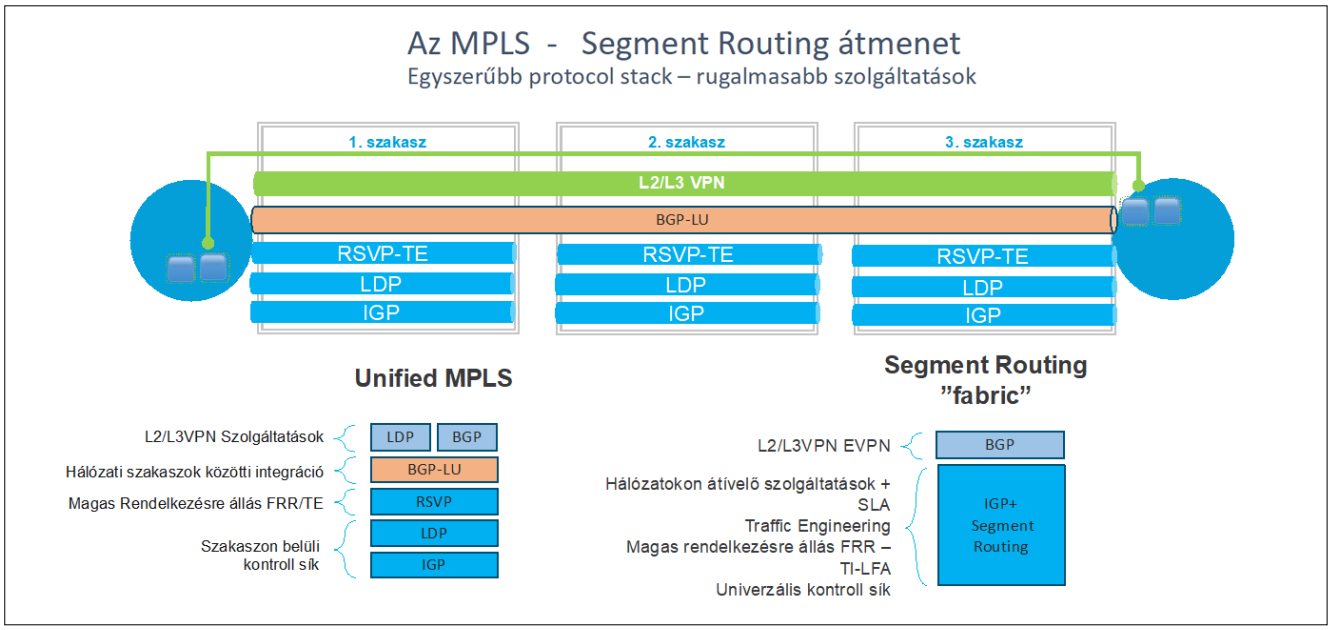
Az MPLS-hálózatokban valamilyen IGP-routingprotokoll (ISIS vagy OSPF) kiszámítja a routingtáblát. Az LDP-protokoll ezekhez címkéket rendel és azokat elküldi a szomszédos routereknek. Ha van Traffic Engineering, az RSVP-protokoll kiszámítja a beállított feltételeknek megfelelő utakat és kialakítja a tunneleket a hálózatban. Ezek tetején ott a BGP, ami a szolgáltatásokért felelős (L2VPN, L3VPN, IPv4, IPv6). Ha nagyon sok hálózati szakaszt kell összekötni, akkor még ott van a BGP-LU (labelled unicast), a skálázhatóság növelésére. Ez a protokollstack látható az ábra bal alsó sarkán.

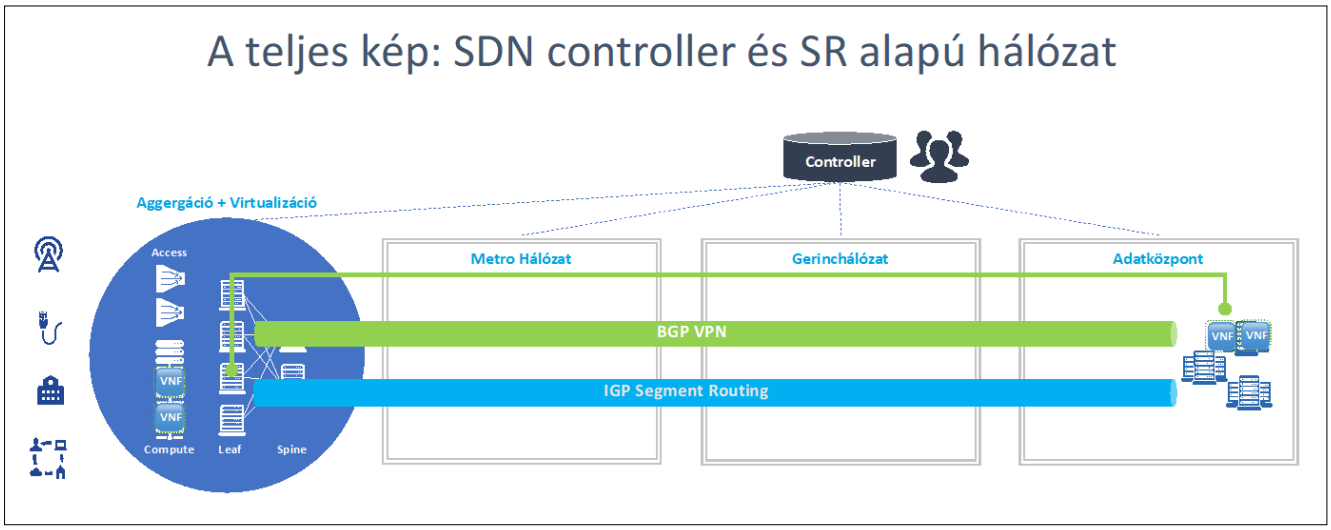
A SR (ábra, jobb alsó sarok) sokkal egyszerűbb! Az IGP-protokoll-kiterjesztések (OSPF és ISIS) magukban hordozzák a SID- (Segment ID) információkat, erről az összes routernek lesz információja a konvergencia befejeződése után. (Többféle SID létezik: prefix SID, adjacency SID, ez lokális, két router közötti link azonosítására szolgál). Ezek után minden MPLS-alapú szolgáltatás működőképes. Ha nagyon gyors konvergencia a követelmény, rendelkezésre áll a TI-LFA (Topology Independent Loop Free Alternate). Ez lényegében azt jelenti, hogy az IGP-protokoll mintegy előre lemodellezi az egyes linkek megszakadását és előre megszerkeszti az ilyenkor használatos label stack-et.

Ha Traffic Engineering szükséges, akkor (legalább) két lehetőség van:

- Az IGP-protokoll különböző paraméterekkel is ki tud számolni utakat (késleltetés, linkek vagy routerek elkerülése stb.). Ezek a label stack-ek előre kiszámíthatók. Ami nagyon fontos, csak az úgynevezett Head End (a kezdő) routernek kell rendelkeznie ezzel az információval, nincs protokoll-interakció a többi routerrel.

3. ábra Az MPLS és Segment Routing technológia összehasonlítása.





4. ábra SDN-alapú, Segment Routing-ot alkalmazó szolgáltatói hálózat.

- A másik megoldás, hogy egy SDN-kontroller az összes hálózati szakaszra kiszámolja a kívánt utat. Ezek után az egyes label stack-eket a megfelelő routerekbe egyszerűen letölti. Ehhez két fontos építőelem szükséges:
  – A kontrollernek el kell küldeni az aktuális hálózati topológiát és a linkek attribútumait. Erre alkalmas – a már az SR-technológia előtt kifejlesztett – BGP LS-(BGP Link State) protokoll.
  – A másik építőelem az SR-szegmensek sorozatának, azaz a label stack-nek a beprogramozása a routerekbe: erre való a PCE- (Path Computation Element) protocol.

Fontos megjegyezni, hogy utóbbi két mechanizmus csak a menedzsment- és kontrollsíkban jelenik meg. A hálózat alapvetően képes üzemelni a kontroller leszakadása esetén is (csökkentett szolgáltatásokkal...).

A fenti leírásnak megfelelő hálózat vázlata látható a *4. ábrán*.

5. Összefoglalás

Cikkünkben megvizsgáltuk a távközlési szolgáltatók hálózati infrastruktúrájának legújabb trendjeit. A növekvő adatátviteli igények és a rugalmasabb szolgáltatásmenedzsment kisebb OPEx- és CAPEx-szintek mellett valósíthatók meg, ha alkalmazzuk ezeket az eszközöket.

- Meg kell vizsgálni az olcsóbb, „merchant silicon” alternatívákat a nagy teljesítményű routerek kiválasztása esetén.
- Meg kell vizsgálni, hogy melyik szolgáltatás virtualizálható, és hogy ezt központi, vagy elosztott architektúrában célszerű-e megtenni?
- Olyan eszközöket célszerű választani, amelyek korszerű API-okkal rendelkeznek és beilleszthetők egy közös orkesztrációs rendszerbe.
- A Segment Routing az MPLS valós alternatívája. A jelenlegi hálózatok átmigrálhatók. A hálózat egyszerűbb lesz és több szolgáltatást képes nyújtani.

Hivatkozások

[1] RFC 8402, Segment Routing Architecture, Clarence Filsfils, Stefano Previdi (eds.), *https://datatracker.ietf.org/doc/html/rfc8402*

A szerzőről

KÁKONYI ISTVÁN 16 éve a Cisco munkatársaként rendszermérnöki, majd architekt pozícióban, túlnyomórészt távközlési szolgáltatókkal foglalkozott. Szakterülete az IP routing/switching, MPLS, távközlési szolgáltatók hálózati architektúrája és az SDN. CCIE és DevNET associate specializációkkal rendelkezik.

Készenléti közszolgáltatás biztosítása járványhelyzetben

CZINTULA GYÖRGY
Pro-M Zrt.
czintula.gyorgy@pro-m.hu

Kulcsszavak: készenléti kommunikáció, közszolgáltatás, felhasználói igény, kihívás válaszai

A cikk bemutatja a több mint 15 éve kormányzati célú hírközlési szolgáltató EDR-közszolgáltatásának fenntartását célzó feladatrendszerében az általános működést meghaladó, járvány okozta felhasználói kihívásokat. A COVID-19 járvány a felhasználóknál megnövekedett többletfeladatokat eredményezett, amik az EDR-szolgáltató részéről válaszokat tettek szükségessé a kihívásokra az EDR-közszolgáltatás zavartalan biztosítása érdekében. Az írás érinti az üzletmenet-folytonosság érdekében meghozott szervezeti, rezsim- és adminisztratív intézkedéseket, valamint összefoglalót ad a járványkezelés eredményességéről, kitekintve a COVID-típusú kihívások jövőbeli kezelésére.

1. Bevezetés

A készenléti szervezetek és a védelemigazgatásban működő szervek 2006 óta a korábbi szegmentált és információbiztonság szempontjából számos kockázattal jellemezhető analóg rádiórendszerek helyett az úgynevezett Egységes Digitális Rádiótávközlő Rendszer (EDR) szolgáltatásait veszik igénybe. Az EDR-rendszer kormányzati célú hírközlési szolgáltatója a Professzionális Mobilrádió (Pro-M) Zrt. Működésének jogszabályi alapja a kormányzati célú hálózatokról rendelkező 346/2010. (XII.28.) Korm. rendelet.

A szolgáltatás mára teljesen beépült a készenléti felhasználók napi szakmai feladataiba. A vezetői, parancsnoki munkát támogató, a felhasználó szervezeten belüli és azok közötti híradótámogatás infokommunikációs eszközévé, a vészhelyzeti kommunikáció nélkülözhetetlen rendszerévé vált. Az általános feladatok mellett hangsúlyosan van jelen havária-helyzetekben, nagy látogatottságú tömegrendezvények, úgynevezett kiemelt biztosítási igénnyel jelentkező események során, a terror elleni védekezéssel összefüggő feladatok és a határmenti migráció kezelésével összefüggő kihívások kapcsán. A készenléti rádiórendszer szerepe jelentősen megnövekedett a SARS-CoV-2 (COVID19) járvány elleni védekezéshez kapcsolódó rendészeti feladatok infokommunikációs támogatásában.

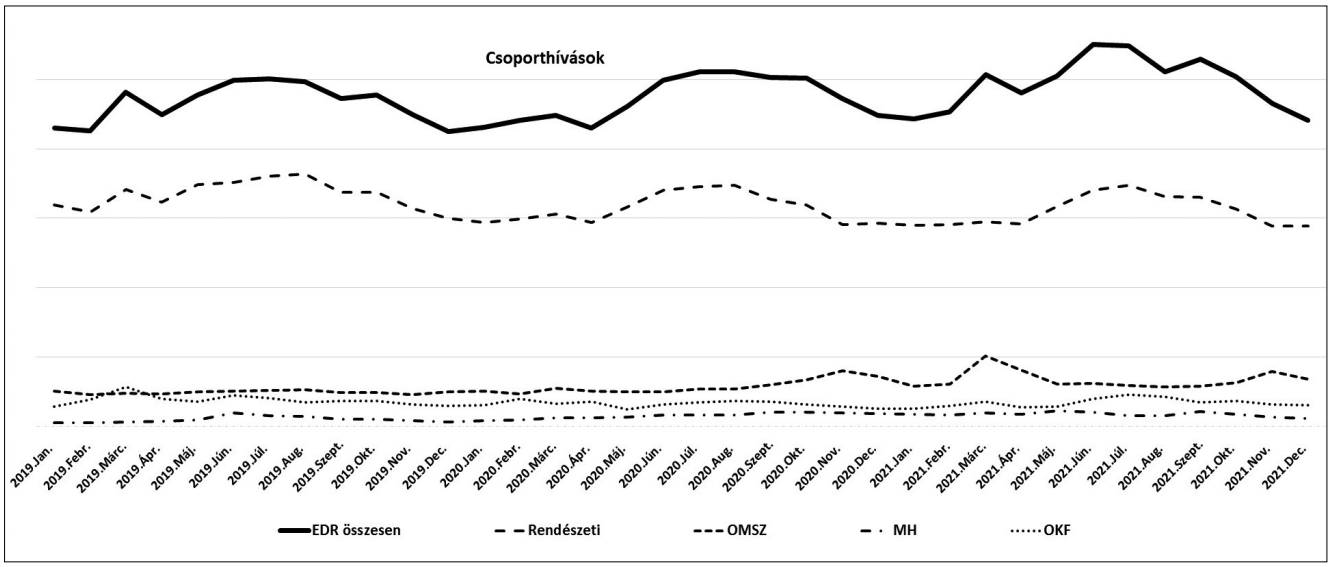
A további szakaszok a készenléti felhasználói körre fókuszálva a járvány már lezajlott, illetve éppen folyamatban lévő hullámainak jelentősen megemelkedett, EDR-érintettségű követelményeit foglalják össze, célul tűzve a készenléti kommunikáció jelentőségének, az EDR-közszolgáltatás fenntartása fontosságának kiemelését. Bemutatják a járvány következtében megjelenő kihívások hatását, a készenléti felhasználók megnövekedett soron kívüli feladataiból eredően jelentkező felhasználói igényeket, valamint az EDR-közszolgáltatás fenntartása érdekében a megnövekedett igényekre adott kormányzati célú hírközlési szolgáltatói válaszokat. Továbbá rövid értékelést adnak a járvány immáron ötödik hullámának időszakában a korábban megtett intézkedések hatásáról, eredményességéről.

2. Kihívások az EDR-hálózattal szemben, a járvány hatása a felhasználó szervezetre

A 2019 decemberében a kínai Vuhan városában megjelent COVID19-fertőzés jelentősen átformálta az elmúlt években mind az egyének, mind a társadalom életét, a nemzeti és világgazdaság működését. Nyomot hagyott a személyiségekben, az államigazgatás, a kormányzat működésén, az informatika, a hírközlés mindennapjain, a közigazgatási informatikai, távközlési rendszerek felhasználásán, és ezek mellett egy speciális területen: az EDR-közszolgáltatást igénybe vevő felhasználók feladataiban, a hálózat, a szolgáltatások és a kormányzati célú hírközlési szolgáltató működésében.

A COVID19-járvány bekövetkezte az általános és kiemelt készenléti felhasználói feladatokon túlmenően jelentkező leterhelést okozott a készenléti felhasználóknak és a kormányzati célú hírközlési szolgáltatónak. A felhasználók és a szakmai felügyeletet ellátó Belügyminisztérium, az anyavállalat részéről egyértelműen megfogalmazódott a járványkezeléssel kapcsolatos felhasználói igények minden körülmények közötti kiszolgálása, a hálózat erőforrásainak és szolgáltatásainak prioritásos biztosítása. Vagyis a hálózat és a szolgáltatások felértékelődtek, elsődleges feladattá vált az EDR-közszolgáltatás biztosítása.

A készenléti felhasználók járványkezelés kapcsán megnövekedett feladatai az EDR- eszközök használatának intenzív emelkedését is jelentették az egyes felhasználó szervezeteken belüli, a szervezetek közötti és a járványkezelésben érintett hatóságokkal, intézményekkel



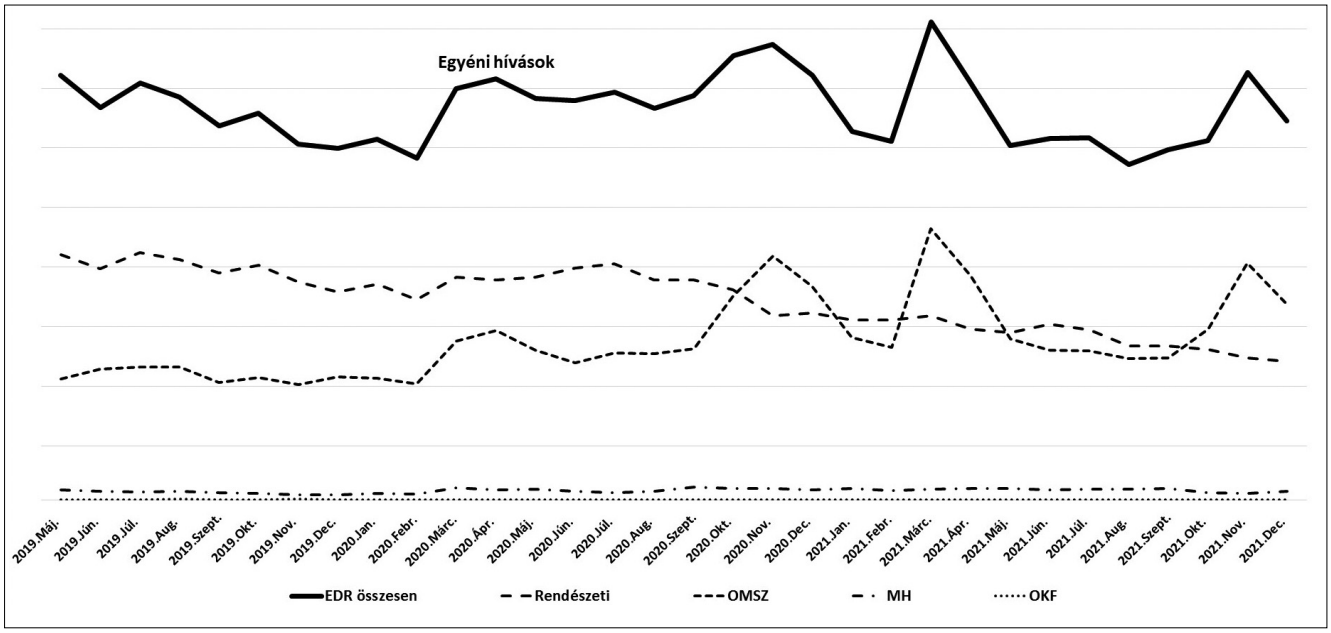
1. ábra Csoporthívások mértékének alakulása az összes hívás viszonylatában.

folytatott rádióforgalmazásban. Különösen érvényes ez a rendészeti, igazgatási feladatok kapcsán az ORFK, továbbá a védekezésben egyedi feladatokat ellátó MH, valamint az OMSZ és az OKF központi, területi szervei esetében *(1. és 2. ábra)*.

Bár a Kormány 2020 januárjában elrendelte a járvány elleni védekezésért felelős Operatív Törzs megalakítását, a járvány első hullámának bekövetkeztekor, a 2020 márciusában kihirdetett veszélyhelyzet időszakában természetesen még nem volt ismert, hogy egy hosszan elnyúló, immáron az ötödik szakaszában járó járványsorozattal állunk szemben, ami a készenléti felhasználók számára is egy hosszantartó rendkívüli időszakot keletkeztetett. A felhasználók és a kormányzati célú hírközlési szolgáltató számára természetesen a COVID19 járvány első hulláma okozta a legnagyobb terhelést 2020 márciusa és májusa között. Amikor a fertőzöttséggel kapcsolatos következmények számszerűen csökkentek, sor került a veszélyhelyzet megszűntetéséről rendelkező 2020. évi LVII. törvény hatályba lépésére. Ezzel egyidőben a 2020. évi LVIII. törvény hatályba lépése járványügyi készültséget rendelt el.

Azonban a rövid átmeneti csökkenésben a megtett intézkedések értékelésére kevés idő jutott, hiszen már 2020 augusztusától bekövetkezett a járvány második hulláma, majd nagyjából 2021 januárjától megkezdődött a COVID19 járvány harmadik hulláma. A harmadik szakasz lecsengése 2021 júniusára tehető, mikor is a járvány elleni védekezésért felelő Operatív Törzs bejelentette a harmadik hullám végét. A napi fertőzöttségi adatok alapján 2021 augusztusában deklarálta az Operatív Törzs a COVID19 járvány negyedik hullámának bekövetkeztét, ami 2022 januárjától az ötödik hullámba csapott át.

2. ábra Egyéni hívások mértékének alakulása az összes hívás viszonylatában.



Az EDR-közszolgáltatás biztosítása érdekében 2020 februárjában a kormányzati célú hírközlési szolgáltató részéről az alábbiak végrehajtására került sor:

– a várható soron kívüli feladatokhoz szükséges készenléti felhasználói rádióterminál- és tartozékigények preventív felmérése;
– a felmérés alapján kapott igények kiszolgálásának megtervezése, ütemezése és feltételeinek megteremtése, a készlet, a gyártói-, beszállítói keretmegállapodások alapján megrendelések, a rendkívüli járványkezelési feladatokhoz kapcsolódó soron kívüli növekvő igények kiszolgálása *(lásd a 3. ábrát)*;
– hatályba lépett a 72/2020. (III. 28.) Korm. rendelet a kórházparancsnokokról és az egészségügyi készlet védelméről, ami a kórházparancsnoki rendszerben érintettek soron kívüli rádióterminállal ellátását igényelte;
– a járvány elleni védekezésben koordináló szerepet betöltő Megyei Védelmi Bizottságok feladatellátásához végre kellett hajtani a működésükhöz szükséges hálózati lefedettség- és kapacitáselemzéseket, eszközellátottságuk mértékének ellenőrzését;
– a Tevékenységirányítási Központok és a felhasználók által alkalmazott szakrendszerek eszközeit érintően biztosítani kellett az elhelyezésből adódó átfertőződés kockázatának csökkentését az eszközrendszerek tartalék vezetési-, irányítási pontokra átcsoportosításával;
– sor került a stratégiai készlet megemelésére, ami az egészségügyi ágazat kijelölt létfontosságú rendszerelemei EDR-hálózathoz történő csatlakozási igényeihez illeszkedő eszközszükséglet tervezését is magában foglalta;
– fel kellett készülni az esetleges – ma már tényként kezelt – további járványhullámok eszközigényeinek biztosítására.

3. A kormányzati célú hírközlési szolgáltató válaszai a kihívásokra, az EDR-közszolgáltatás fenntartására hozott intézkedései

3.1. Adminisztratív intézkedések

A Kormány 2020 januárjában rendelte el a járvány elleni védekezésért felelős Operatív Törzs megalakítását, továbbá kihirdette a veszélyhelyzetet a 40/2020. (III. 11.) Korm. rendelet hatályba lépésével.

A kormányzati célú hírközlési szolgáltató rendelkezésre állása szempontjából fontos volt, hogy a Pro-M Zrt. szervezetében a vezérigazgató vezetésével már 2020. február 11-én megalakult a Pandémiás Tanács, a Társaság Működésfolytonossági és katasztrófaelhárítási tervében foglaltak alapján, mint a rendkívüli helyzetek kezelésére hivatott válságstáb. A válságstáb/Pandémiás Tanács elsődleges feladata a kialakult rendkívüli helyzet kockázati hatásainak minimalizálása; a veszélyhelyzet-kezelés szakaszaiban elvégzendő feladatok meghatározása és a szükséges feltételrendszer biztosítása; a kritikus munkakörökhöz/folyamatokhoz kapcsolódóan a feladatkörök, a felelősségek és a jogosultságok meghatározása; valamint az együttműködés rendjének kialakítása.

Az első lépések egyikeként sor került a már 2008-ban kidolgozott és hatályba léptetett, évente felülvizsgált Működésfolytonossági és katasztrófaelhárítási terv COVID19-járvány elleni védekezés okán szükségessé vált felülvizsgálatára. Mindennek megfelelően sor került

3. ábra A felhasználói eszközigény alakulása.

COVID-19 első hullám eszközigény	
Felhasználó	**db**
Rendőrség	936
HM	600
OMSZ	200
Kórházparancsnokok	126
Kijelölt egészségügyi ágazati létfontosságú intézmények	340
Pro-M munkatársi kör ellátása	70
Összesen	**2272**

COVID-19 második harmadik és negyedik hullám eszközigény	
Felhasználó	**db**
Rendőrség	456
HM	300
OMSZ	370
Kórházparancsnokok	126
NAV	31
Magyar Vöröskereszt	10
Kijelölt egészségügyi ágazati létfontosságú intézmények	340
Pro-M munkatársi kör ellátása	92
Összesen	**1725**

Központi stratégiai rádióterminál tartalékkészlet
Döntően azóta is a felhasználóknál rendelkezésre áll, illetve a készletben elkülönített
(a hálózatban használt eszközszám 5%-a)

a rendkívüli helyzetben, a telephelyen ellátandó kritikus munkakörök és létszám, helyettesítő munkatársi kör áttekintésére. A kormányzati járványkezelési intézkedések és az EDR-közszolgáltatás minden körülmények közötti fenntartása szükségessé tette a járványügyi helyzetekre való felkészülés szabályait meghatározó dedikált tematikus belső szabályzat létrehozását (járványügyi szabályzat).

Ennek célja a Működésfolytonossági és katasztrófaelhárítási tervben rögzítettekkel összhangban a járványügyi helyzetek kapcsán szükséges intézkedések, szabályok és feladatok részletes meghatározása a Pro-M Zrt. minden szervezeti egységére és dolgozójára vonatkozó érvényességgel. Meghatározza a Pandémiás Tanács hatáskörét és feladatait, továbbá lebontja a Társaság vezetőjére, a szakmai területek vezetőire és a munkatársakra érvényes magatartási szabályokat, kötelező feladatokat, rögzíti az otthoni munkavégzés szabályait.

Megtörtént az otthoni munkavégzés szabályzatának felülvizsgálata és a védekezési feladatokhoz igazítása. Hangsúlyosan kiemelendő, hogy az otthoni munkavégzés nem vonatkozik az ügyeleti területeken dolgozó munkatársakra, hiszen ezeknek az egyedi feladatoknak a feltételrendszere alapvetően csak a Pro-M Zrt. telephelyén biztosított teljes körűen.

Egy havária-helyzetben, illetve a működésfolytonosságot veszélyeztető esetekben megvan a lehetőség más, kijelölt és előre felkészített telephelyen az ügyeleti feladatrendszer ellátására csökkentett funkcionalitással, ugyanakkor ezen feladatoknak az otthoni munkavégzés keretében ellátása a biztonsági követelmények hiánya okán nem lehetséges, illetve sem az iparágban, sem a készenléti felhasználók körében nem alkalmazott modell.

3.2. Intézkedések a működésfolytonossághoz

Vállalati erőforrások felmérése, biztosítása	A járvány során a kialakuló helyzetnek megfelelően a kritikus munkakörök és kritikus létszám felmérése; helyettesítési, átcsoportosítási lehetőségek számbavétele – különösen a kritikus munkakörökben és az ügyeleti rendszerben – az erőforrások rendelkezésre állása; az esetleges nagyarányú HomeOffice elrendelése, munkába járási nehézségek kezeléséhez szükséges erőforrások biztosítása.
Üzemeltetési peremfeltételek meghatározása	Az EDR hálózati infrastruktúra műszaki üzemeltetése és a szolgáltatás zavartalan működésének biztosítása; üzemeltetési kockázatok minimalizálása; saját és alvállalkozói rádiós és kapcsolóközponti tartalékanyag-készletek felmérése, igény esetén átcsoportosítása; vészhelyzeti kommunikáció biztosításában érintett IT-rendszerek felügyelete; felhasználói igény esetén rádióprogramozások biztosítása.
Fejlesztések felfüggesztése	A hálózati erőforrások a felhasználók számára folyamatosan és zavartalanul rendelkezésre állása érdekében az éles hálózatban tervezett fejlesztések ideiglenes felfüggesztése, szüneteltetése; ez egyben a fejlesztésekben közreműködő külső erőforrások távolmaradását is jelenti, ami az esetleges fertőzési kockázatok csökkentését, a belső erőforrások kímélését munkaszervezési és együttműködési korlátozások bevezetését is jelenti.
Szolgáltatói változtatások befagyasztása	A fejlesztések ideiglenes szüneteltetése, felfüggesztése; egyidejűleg a hálózatban, a szolgáltatásokban végrehajtandó változtatások, módosítások (szoftver- és hardverváltoztatások, upgrade-k) tilalma.
Külső partnerek erőforrásainak felmérése	Az egyes szakterületek – különösen a kritikus munkakörök, ügyeleti egységek – tekintetében felkészülés az esetleges átfertőződések okán előforduló erőforrás-kiesésekre; a helyettesítésbe, átcsoportosításba bevonható belső és külső – például rendszerszállító – résztvevők felmérése, (szakmai, helyismereti) felkészítése.
Üzemeltetői létszám biztosítása	A kritikus üzleti folyamatok kiszolgálásához szükséges létszám folyamatos biztosításához a munkarend/szabadságolások/távollétek áttekintése, esetleges ideiglenes módosítása.
Ügyeleti rend áttekintése	Az erőforrás-átcsoportosításra, a védekezésre kijelölt Operatív Törzs támogatására az ügyeleti rend átfogó áttekintése; az ügyeletek szükség szerinti funkcionális átrendezése, összevonási lehetőségek vizsgálata, rendkívüli helyzetekhez kapcsolódó ügyeleti feladatokkal való felruházása.
Ügyeletek távoli munkavégzésének előkészítése	A Társaságon belüli magas átfertőződés esetére, vagy éppen az azzal kapcsolatos kockázatok hatékony és jelentős csökkentése érdekében a megszokottól eltérő ügyeleti feladatvégzés modelljére áttérés feltételeinek előkészítése, a szükséges jogosultságok biztosítása.
Ügyeleti munkahelyek szeparálása	Az általánostól eltérő ügyeleti működési modell bevezetése, az ügyeletek átfertőződési kockázatának csökkentése érdekében az egyes ügyeleti területeken az ügyeleti munkahelyek fizikailag eltérő elkülönítéséről, szeparálásáról gondoskodás, a szükséges munka- és rendszerfeltételek kialakítása.
Áttérés az EDR-eszközök használatára	A napi működésben a Társaság vezetői és munkatársai vészhelyzeti elérhetősége és biztonságos kommunikációja érdekében a társasági szintű EDR-kommunikáció bevezetése.
Védekezésre szolgáló eszközök biztosítása	Tisztítószerek, fertőtlenítő szerek készletezésének előkészítése, figyelemmel az esetleges áruhiányra és meghosszabbodó szállítási időkre.

Munkatársi tájékoztatás	A munkatársak folyamatos, hiteles és kiegyensúlyozott tájékoztatása heti rendszerességgel, hírlevélben és az Intraneten keresztül megvalósuló tájékoztatással.
Felkészülés a fokozott, rendszeres információcserére	A belső információs tájékoztatás mellett külön figyelem a gyártó, beszállító, együttműködő partnerekkel, ügyfelekkel, felügyeleti szervvel, tulajdonos szervezettel történő folyamatos kapcsolattartásra, az információk felfrissítésére a fokozott, célirányos pandémiás információcserére.
Válságstáb/Pandémiás Tanács összehívása	A vonatkozó belső szabályzat alapján, szükség esetén vezérigazgatói döntésre a Válságstáb/Pandémiás Tanács rendszeres ülésein a helyzet értékelése, a védelmi intézkedések hatásának elemzése, a járvány alakulásának függvényében további indokolt intézkedések meghozatala.
Operatív Törzs közvetlen támogatása (igény esetén)	Ügyeletek munkarendjének átalakítása és igény szerint EDR-szakértő delegálása, felkészülve és biztosítva a védekezés irányításáért kijelölt Operatív Törzs zavartalan munkáját 24 órás, folyamatos helyszíni jelenléttel.
A létfontosságú magyar vállalatok biztonságáért felelős akciócsoporttal együttműködés	Az EDR-hálózat szerepére, a közszolgáltatás fontosságára és felhasználói köre minőségére figyelemmel az Operatív Törzs irányú együttműködés és kapcsolattartás segítése.
Együttműködés kiterjesztése a felügyeleti szervekkel, hatóságokkal	A Belügyminisztériummal megkötött Közszolgáltatási Szerződésben meghatározottak szerint a „vis major" helyzet kezelésére, az együttműködő érintettek tájékoztatására, illetve közös projekt teamek felállítására való felkészülés a Tulajdonos képviselőjével, a felügyeleti hatóságokkal (OKF, NMHH) a pandémia okozta károk enyhítése, a védekezés hatásainak, kockázatainak kezelése.
Szabadságolások, távollétek felfüggesztése, szabadságok beszüntetése	Amennyiben a járvány alakulása, illetve a meghozott védelmi intézkedések indokolttá teszik, az EDR-közszolgáltatás biztosítása érdekében a szabadságolások és egyéb okok miatti távollétek felfüggesztése/korlátozása, az üzemeltetéshez szükséges létszám biztosítása.
Tömegközlekedés korlátozása	A munkatársak megfertőződése, további átfertőződése, a fertőzöttség kockázatainak csökkentése – a lehetőségek keretei között – a tömegközlekedési eszközök igénybevételének korlátozása, tiltása, a vállalati gépkocsipark és magán gépjárművek használatának elrendelése, telekocsi-rendszer szervezése.
Pandémiás higiénia biztosítása, kézmosó pontok pandémiás megerősítése	A védekezésben és a fertőzések megelőzésében a kézfertőtlenítő állomások kihelyezése a belépési pontokra és azok folyamatos üzemeltetése, továbbá célzottan az adott helyzet kezelésére alkalmas tisztítószerek biztosítása (pl. antibakteriális kézmosók).
Emelt szintű takarítás	A higiéniai, szellőző- és klímarendszerek működtetését, a takarítást, fertőtlenítést, hulladékszállítást, veszélyes hulladék kezelését célzó szigorítás; az ügyeleti munkahelyiségek, az ott használt eszközök – különösen, ha azokat többen kezelik –, a mellékhelyiségek, közösségi terek, gyakran használt eszközök (kilincsek, kapcsolók, biztonsági tasztatúra, stb.) szokottnál gyakoribb fertőtlenítőszeres tisztítása.
Pandémiás tájékoztatók kihelyezése	A helyzet kezelését támogató tájékoztatók, ismertetők kihelyezése, célzott kommunikálás gyakorlata.
Külső, belső oktatások, megbeszélések online térbe irányítása	Oktatások, képzések, megbeszélések, konferenciák, vendégek és látogatók fogadásának szüneteltetése, online térbe irányítása.
Külföldi magán és hivatalos utazások követése, szükség esetén korlátozása	Külföldre történő kiutazási szándékról legkésőbb a kiutazást megelőző második munkanapon kötelező a közvetlen vezető írásbeli (elektronikus úton történő) tájékoztatása, hivatalos kiutazások eseti engedélyeztetési kötelmének bevezetése.
Vendégek érkezésekor kötelező figyelemfelhívás	Az irodaépületek belépési pontjain a vendégek érkezésekor kötelező figyelemfelhívás (írásban, szóban) a védőeszközök használatára (kézfertőtlenítő, szájmaszk, kesztyű, stb.)
Pandémiás nyilatkozat bevezetése	Az irodaépületek belépési pontjain pandémiás adatlap felvétele, egyben pozitív nyilatkozat esetén a belépés tiltása.
Tünetvizsgálat elrendelése	Az irodaépületek belépési pontjain pandémiás tünetvizsgálat bevezetése (pl. testhőmérsékletmérés), amely eredményétől függően a belépés megtiltása.
Kötelező otthoni munkavégzés elrendelése	A lehetséges munkakörökben a kötelező hosszú idejű, meghatározott idejű otthoni munkavégzés bevezetése.
Fokozott (hosszabb időtartamú) otthoni munkavégzés elrendelése	Tervezetten a kialakult helyzethez igazodva a szakmai feladatokra figyelemmel az „egy iroda – egy személy" munkarend kialakítása.
Elektronikus iratforgalom, digitális aláírás bevezetése	A személyes jelenlét elkerülésével az online térben végzett munka digitális feltételeinek megteremtése.
Munkatársak védettségi szintjének figyelemmel kísérése	Kritikus infrastruktúra-elem üzemeltetőként kiemelten fontos az oltottság elsődlegessége, az ellenanyagszint-mérés lehetőségének biztosítása.
Védőeszköz biztosítása	A kellő védettség eléréséhez, a fertőzés kialakulásának és az átfertőződöttség megelőzéséhez szükséges eszközkészlet kiadása minden munkavállalónak.
Ózonos helyiség-fertőtlenítés alkalmazása	Az ügyeleti helyiségek – és azon munkaszobák, melyek esetében a munkatársakat érintő fertőzések fordultak elő –, rendszeres ózonos fertőtlenítése.

3.3. Rezsim-intézkedések a járvány kockázatainak csökkentésére

4. COVID-típusú kihívások kezelése a jövőben (a jövőben?)

• A COVID19-járvány 1- 4. hullámának tapasztalatai alapján a preventív és proaktív reagálásnak kiemelt jelentősége van a járványkezelés eredményessége érdekében.

• Az eszközkészletek naprakészen tartása a rövid idő alatt megemelkedő eszközigények kiszolgálását biztosítja.

• A készletek feltöltésénél figyelemmel kell lenni a járvány korlátozásai okozta, a beszállítói/gyártói körben a rádióterminál ellátásban várhatóan jelentkező csúszásokra, gyártói, beszállítói alkatrészellátási nehézségekre (chip-ellátási problémák stb.).

• A létfontosságú rendszerelemek EDR-eszközzel ellátásának a készenléti felhasználók igényei kiszolgálásával egyidejűleg történő kezelése.

• Az általánostól eltérő hálózat- és rádióterminál-használat a forgalomszervezés és erőforrás-menedzsment felülvizsgálatát teszi indokolttá.

• A készenléti felhasználók megemelkedett és megváltozott hálózat- és rádióterminál-használata a későbbi fejlesztésekben a szélessávú applikációkban a felhasználói igények megvalósítását indokolja.

• A tapasztalatok értékelése alapján a kormányzati célú hírközlési szolgáltató

részéről elvégzett feladatokat célszerű beépíteni a Működésfolytonossági és katasztrófaelhárítási tervhez kapcsolódó éves BCP/DRP-gyakorlatokba.

• A járvány elleni védekezés alatt bevezetett intézkedések adott időszakot követő kivezetése az indokolt intézkedések további megtartását nem teszi szükségtelenné.

• Különösen a Társaságon belüli folyamatos kommunikáció és összetartozás erősítése a bevezetett korlátozások komplex hatásának, káros következményeinek csökkentését biztosítja.

5. Az intézkedések értékelése, megtartása, továbbfejlesztése

• A járvány elleni védekezéshez kapcsolódó oltakozás érdekében megalkotott Oltási Terv változásainak és prioritásainak módosulásai okán Társaságunk a munkavállalói részére gyorstesztelés, ellenanyagszint mérés lehetőségét biztosította.

• Az egyéni döntés alapján felvehető első, majd ismétlő oltások hatására kialakuló védettség vált a pandémiás helyzet kezelésének legfőbb eszközévé, egyben jelen pillanatban is a fertőzés megelőzésének prioritásos eszköze.

• A védettséget kialakultnak tekinthettük az első oltást követő második hét végétől a munkatársak több, mint 90%-át érintően.

• A megfelelő átoltottság és a bevezetett intézkedések hatására kialakult, illetve fenntartott állapot, valamint a COVID19 hullámának – ma már tudjuk – átmeneti lecsengése lehetővé tette az intézkedések enyhítését, a fokozatos nyitás végrehajtását, ami 2021. május 21-től kezdődött meg.

• Erre alapozva 2021. június 11-én a Pandémiás Tanács ideiglenesen felfüggesztette munkáját.

• Az alapvető védelmi intézkedések és bevezetett eljárások megtartása mellett a COVID19-járvány 4. hullámának alakulására figyelemmel 2021. szeptember 27-én a Pandémiás Tanács újra összeült, és azóta is irányítja a védekezési feladatokat heti rendszerességgel megtartott ülésein és döntésein keresztül.

• Az online tér használatának lehetősége továbbra is nyitva állt, illetve a személyes jelenlétű megbeszélések mellett jelentős mértékű maradt.

• Az otthoni munkavégzés kiterjedtségére figyelemmel az elektronikus iratforgalom, a digitális aláírás eszköze a továbbiakban is prioritást kap.

• Annak érdekében, hogy a Társaság munkavállalói védettségének mértéke ismert legyen, ellenanyagszint mérés lehetőségét biztosítottuk.

• Figyelemfelhívás, egyéni felelősség, védekezési eszköztár használata.

6. Összefoglalás

A cikk összefoglalta a COVID19-járvány 2020. januári magyarországi megjelenése óta eltelt időszakban az EDR-közszolgáltatással szembeni, havária-körülmények közötti ugrásszerűen megnövekedett felhasználói igényeket, az igények által generált kormányzati célú hírközlési szolgáltatói válaszokat. Bemutatta a járvány elleni védekezés adminisztratív, üzletmenet-folytonossági és rezsim-intézkedéseit, egyben értékelte azok jelenlegi és jövőbeni hatásosságát.

A sokrétű felhasználói kapcsolatrendszerben jelentős szerepet betöltő felsővezetői értékelő találkozókon elhangzottak megerősítették, hogy a készenléti felhasználók megnövekedett soron kívüli járványkezelési feladatait az EDR-rendszer kiszolgálta, a kormányzati célú hírközlési szolgáltató munkatársi köre biztosította, azokat jelenleg is kiszolgálja és biztosítja. A kormányzati célú hírközlési szolgáltató munkavállalói körében kisszámú megbetegedések történtek, súlyos, a működésfolytonosságot veszélyeztető erőforrás-kiesések nem fordultak elő.

A szerzőről

CZINTULA GYÖRGY 1978–1997 között a Belügyminisztériumban dolgozott. Egyetemi diplomát 1991-ben a Zrínyi Miklós Katonai Akadémia Rendszerszervező és Vezetés Automatizálási tanszéken szerzett. Egyetemi doktori címét 1993-ban szerezte, értekezésének címe: „Az adatvédelem aspektusainak elméleti kérdései és gyakorlati megvalósulása a belügyi számítástechnikai rendszerekben”. 1997-től dolgozott a távközlés területén, a Westel 900 GSM Mobil Távközlési Rt.-nél, illetve utódszervezeteinél, majd a Magyar Telekomnál. Elsősorban információbiztonsággal, információvédelemmel foglalkozott. Jelenleg a készenléti felhasználók zártcélú hálózatának kormányzati célú hírközlési szolgáltatója, a Pro-M Zrt. ügyfélkapcsolati igazgatója. Alelnöke a Promotel Professzionális Távközlésért Magyországi Egyesületnek, tagja a HTE-nek, a Hétpecsét Információbiztonsági Egyesületnek, valamint a Magyarországi Biztonsági Vezetők Egyesületének.

Az edge computing mint disztruptív technológia

KOVÁCS BENEDEK
Ericsson Magyarország Kft.
benedek.kovacs@ericsson.com

Kulcsszavak: edge computing, 5G, IoT, cloud computing, industry 4.0, mobilhálózati szolgáltatások

Az edge computing napjaink egyik legnépszerűbb technológiája, több neves piackutató is évről évre beválasztja a legfontosabb trendek közé. Bár a koncepció évek óta létezik, az 5G- és IoT-rendszerek megjelenésével került ismét előtérbe.
A cikkben azt fogjuk bemutatni, hogy miért tekinthetjük disztruptív technológiának.
Rövid bevezetés után áttekintjük az edge computing legfőbb felhasználási eseteit, áttérünk a különböző iparági szereplőkre, motivációikra és lehetséges értékláncokat mutatunk be. A harmadik szakaszban technológiai megoldásokat tárgyalunk, illetve néhány megoldatlan problémát vázolunk fel.

1. Bevezetés

Az edge computing koncepcióját az internetes alkalmazások elterjedésével egy speciális típusú kliens-szerver alkalmazás architektúrára kezdték el bevezetni. Ezekben az architektúrákban szervereken futó feladatokat helyezünk a kliens alkalmazáshoz közelebb, földrajzi, hálózattopológiai vagy egyéb, a felhasználói alkalmazás által megkövetelt szempontból. Pontos definícióról nem beszélhetünk, inkább azt érdemes megvizsgálni, hogy mi motiválja ma ezt az új típusú architektúrát a felhőalapú számítások és az 5G-hálózatok korában.

Az edge computing egyértelműen a 3GPP által szabványosított 5G-vel került előtérbe. Az 5G-szabványok legfontosabb elemei a 4G-hez képest jóval kisebb késleltetésű és nagyobb sávszélességű rádió, valamint a szolgáltatásközpontú architektúrában megtervezett 5G-maghálózat. Az edge computing szempontjából kulcsfontosságú a helyi kicsatolás, amely funkció már a 4G-hálózatokban is jelen volt, ahol is a gerinchálózati forgalom optimalizálására használták. Az 5G kis késleltetésű rádiójának sok esetben kötelező kiegészítője a helyben kicsatolt felhasználói forgalom, amely nélkül az 5G tipikus felhasználási eseteinek követelményei nem teljesülnének.

2. Az edge computing felhasználási esetei és a belőlük származtatott követelmények

Az 5G a kezdetektől fogva nem csak a fogyasztói szegmenst célozta meg, hanem a vállalatit is. Olyannyira, hogy míg a fogyasztói szegmens esetén igazából a felhasználó számára szinte észrevehetetlen kiegészítése a 4G-hálózatnak, a vállalati szegmensben az 5G és a vele járó üzleti folyamatok ténylegesen új felhasználási módokat tesznek lehetővé.

A fogyasztói szegmensben fontos újdonság, hogy az 5G-rádió nem csak letöltésre (downlink) hanem feltöltésre (uplink) is optimalizálható, méghozzá programozható módon. Ennek eredményeképpen, illetve új fogyasztói eszközök megjelenésével lényegesen nőtt az internet forgalma. Kifejezetten a video- és egyéb médiaforgalom agresszív növekedését figyelhetjük meg: „A mobilhálózatok majdnem 300-szor akkora forgalmat szolgálnak ki, mint 2011-ben.” [1]

Az adatforgalom mennyisége azonban nem minden. Jó néhány új típusú, a mobil hálózatokon keresztül elérhető, vagy majdan elérhető médiaalkalmazás nem csak a sávszélesség, de a késleltetés szempontjából is kritikus. Egy XR-alkalmazás esetén (XR, azaz mixed reality, azaz a virtuális valóság és a kiterjesztett valóság kombinációja), a kamera mozgása és a tartalom megjelenítése közötti eltelt idő nagyban meghatározza a felhasználói élményt. Ugyanilyen lényeges az interakciók kezelése, például két felhasználó interakciója az ilyen alkalmazásokban. Több tanulmány kimutatta, hogy még ha a hálózatra kapcsolt eszköz képes is nagy komplexitású számítások futtatására, például képalkotásra (angolul rendering) vagy képfelismerésre mesterséges intelligencia segítségével, ám ezen alkalmazások erőforrás-igénye hamar lemeríti az eszköz akkumulátorát. Ezekben az esetekben a számítások egy része elvégezhető a felhőben, de ahhoz, hogy egy tárgyfelismerés megfelelően gyorsan végbemenjen, a felhőszerver és az alkalmazás-kliens közötti késleltetésnek alacsonynak kell lennie [2].

Az 5G-hálózatok, illetve az általuk ígért alacsony késleltetés és megbízhatóság lehetővé teszik sok, úgynevezett vertikális iparág modernizációját. Az egyik példa erre a mobilitási iparág, azaz az autóipar és a közlekedés. Az autóipar legújabb és tervezett mobilitási szolgáltatásai hatalmas adatforgalom-igénnyel rendelkeznek. 2025-re az előrejelzések szerint kb. 100 petabyte–10 exabyte-ot fog kitenni az autók által forgalmazott adatmennyiség. Ezt persze előre nem tudjuk, de az autógyártók

és telekommunikációs iparági szereplők által létrehozott Automotive Edge Computing Consortium-ban jelenleg is dolgoznak egy elosztott architektúrán, amely támogatja az új típusú mobilitási alkalmazások fel- és letöltési adatforgalom-igényét.

A negyedik ipari forradalom is sok új felhasználási esetet definiál az 5G-hálózatok bevezetésével. Ezen esetek közül sok igényli az alacsony késleltetést és a magas rendelkezésreállást. Egy másik részük viszont akár 4G-rádiós hozzáféréssel is kiszolgálható, mert a legfontosabb igény a megbízható rádiós kommunikáció, adat- és hálózatbiztonság. A 3GPP definiál úgynevezett nem publikus hálózatokat (non-public networks), melyek kifejezetten a cégek, kormányzati és egyéb magánhálózatok speciális igényeit szolgálják ki. Már több olyan teszthálózat létezik, ami helyi lefedettséget biztosít egy-egy gyártelepen, kikötőben, vagy egyéb ipari területen.

Ezek alapján a következő szempontok motiválják az edge computingot technikai oldalról: a gerinchálózati sávszélesség optimalizációja, a felhasználói eszköz energiahatékonyságának növelése (nagy számításigényű alkalmazások felhőbe vitelével), az alacsony késleltetés elérése, biztonság, kiváltképpen az adatbiztonság.

3. Új üzleti kapcsolatok a hálózati és felhő-infrastruktúra kiépítésében

A technológiai motivációkon és követelményeken kívül az edge computing új értékláncokat vezet be a telekommunikációs szolgáltatók vállalati üzletágában. A telekommunikációs ipar szereplőin kívül kulcsfontosságú a felhőszolgáltatók, az IT-cégek és az alkalmazást fejlesztő vállalatok szerepe.

A jelenlegi mobilalkalmazások tipikus architektúrájára és a mögöttük álló értékláncra adunk egyszerűsített példákat ebben a szakaszban. Az alkalmazások tipikusan egy felhőben futó szerveralkalmazásból és a felhasználói eszközökön futó kliensalkalmazásból állnak. Az alkalmazásfejlesztők abban érdekeltek, hogy a hálózati szolgáltatótól függetlenül azonos felhasználói élményt nyújtsanak a világon bárhol (illetve ipari alkalmazások esetén egy adott helyszínen).

Az 5G-maghálózatot szolgáltatásorientált architektúrával tervezték, ami praktikusan azt jelenti, hogy a hálózati funkciókat (routerek, kapcsoló logikák, adatbázisok) úgynevezett Cloud Native módon fejlesztik [3]. A hálózati funkciók emiatt tetszőleges felhő-infrastruktúrára telepíthetőek [4].

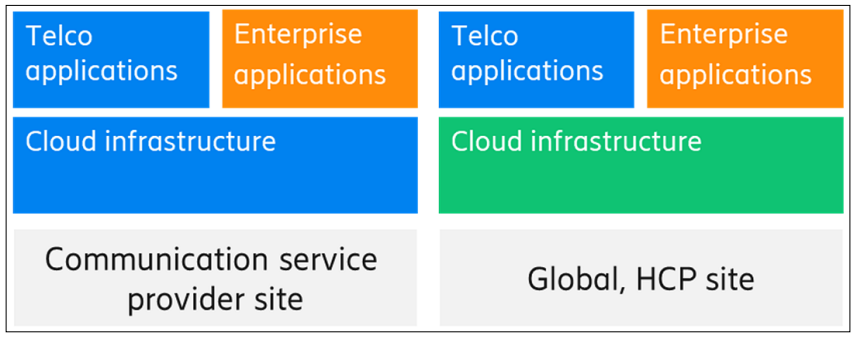
A továbbiakban több lehetőséget is bemutatunk, hogy az infrastruktúra-alkalmazás felosztásban milyen kivitelezési opciókat találunk az egyes telekommunikációs szolgáltatók, IT-infrastruktúra- és felhőszolgáltatók, illetve az alkalmazást fejlesztők (vállalatok), mint piaci szereplők között.

Az *1. ábra* bal oldalán a jelenleg is alkalmazott telepítési opció látható, ahol a kékkel jelzett mobilszolgáltató üzemelteti mind a felhő-infrastruktúrát, mind a mobilhálózatot. A jobb oldalon megjelenik annak lehetősége, hogy telekommunikációs funkciókat (kék) telepítsünk globális felhőszolgáltatók (zöld) infrastruktúrájára. Narancssárga színnel a vállalati ügyfél által fejlesztett és üzemeltetett funkciókat jelöltük. Fontos megjegyezni, hogy a felhasználók számára a vállalat nyújtja a szolgáltatást az alkalmazáson keresztül, a mobil- és felhőszolgáltatók ilyen értelemben a kommunikáció, kapcsolódás, mobilitás és az alkalmazásfuttatási környezethez szükséges komponenseket nyújtják.

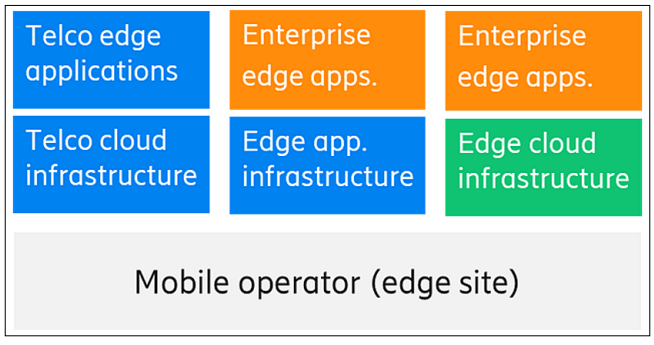
A *2. ábrán* egy hálózatperemi (edge) kivitelezést mutatjuk be. A telekommunikációs funkciókat jelenleg is speciális követelményeknek és az ETSI NFV-szabványnak megfelelő telekommunikációs infrastruktúrára telepítik, mellyel biztosított nem csak a szabványos kompatibilitás, hanem a magas rendelkezésreállás és hatékonyság is. A felhasználói és vállalati alkalmazások helyet kaphatnak az IT-szolgáltatók által telepített felhő-infrastruktúrán, vagy a telekommunikációs szolgáltatók által telepített, speciális edge-felhőn.

A *3. ábrán* egy speciális, úgynevezett nem publikus hálózati megvalósítást láthatunk. Egy ilyen kísérletet mutatott be a Telefónica német vállalata, az Amazon Web Services-szel és az Ericssonnal közösen [5].

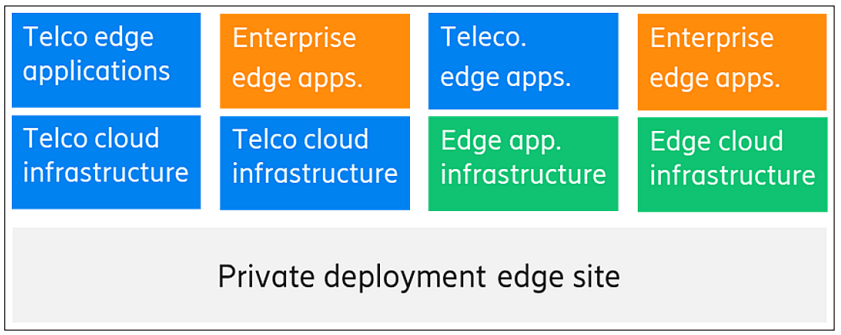
Egyfelől, a globális felhőszolgáltatók üzletet látnak kisebb, specifikus adatközpontok létrehozására a hálózat peremén, és bizonyos esetekben mobilszolgáltatásokkal kombinálják ezt. Másfelől pedig, a telekommunikációs szolgáltatók az 5G-hálózatokat úgy alakítják ki, hogy azok platformként szolgáljanak az alkalmazások számára testre szabott szolgáltatásminőséggel és programozhatósággal.



1. ábra Architektúra-opciók a hálózat adatközpontjaiban.



2. ábra Architektúra-opciók a hálózat peremén.



3. ábra Privát edge computing megoldás lehetséges architektúrája.

4. Megvalósítás

A 3GPP-szabvány több elemet is tartalmaz az edge computing támogatására, de a telekommunikációs iparág más fórumainak, például az ETSI-nek és a GSMA-nak is vannak ide vonatkozó ajánlásai. Sok más iparági összefogás, nyílt forráskódú és de facto szabvány tesz javaslatot a hálózati architektúra és az alkalmazási platformok kialakítására.

Az Automotive Edge Computing Consortium és az 5G Automotive Association teljes megoldásokat dolgoz ki a hálózatba kapcsolt járművek támogatására. Az 5G Alliance for Connected Industries and Automation pedig az ipari felhasználási eseteket tárgyalja és ajánlásokat tesz az ipar 4.0 követelményeinek kielégítésére. Az IT-szolgáltatók, ezen belül is a globális felhőszolgáltatók több szinten állnak elő megoldásokkal, kezdve az integrálható IoT-gateway-ektől a teljeskörű, helyben telepíthető megoldásokig.

A jelen cikkben nem szabványok és ipari megoldások szerint, hanem megoldandó problémák szerint vesszük végig azokat a funkciókat, melyek vagy szükségesek az edge computing-hoz, vagy támogatják azt. A lista természetesen nem teljes, mint ahogy az edge computingnak sincsen pontos definíciója.

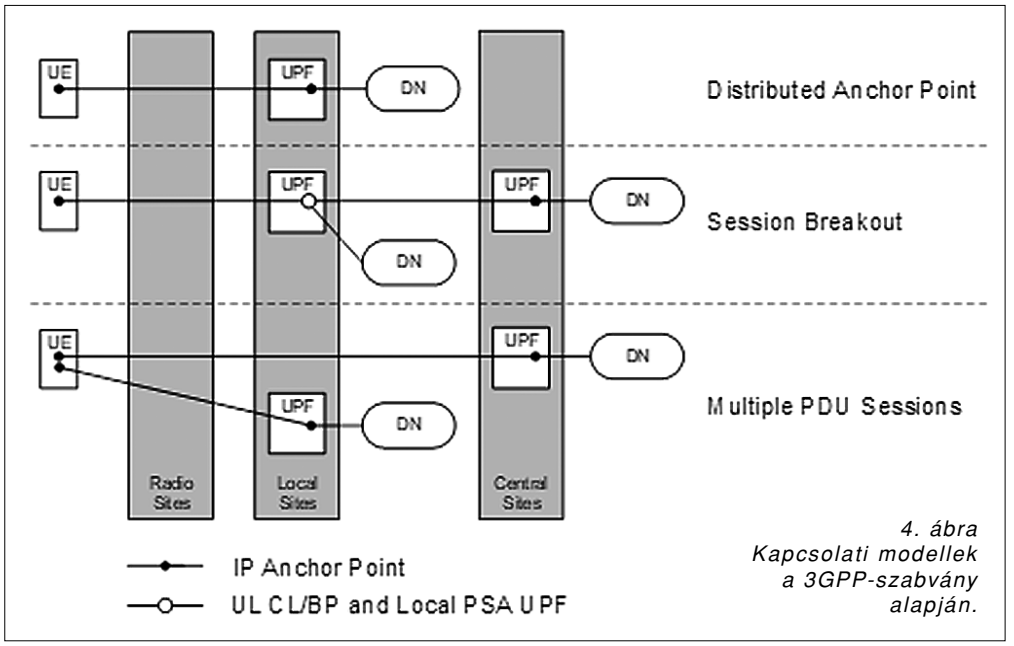
4.1. Az adatforgalom irányítása

A 3GPP-szabvány az edge computing architecktúra alapjait a helyi kicsatolás (local breakout) bevezetésétől támogatja. Megjegyzendő, hogy a local breakout kifejezést a 3GPP-ben már nem használják a következő kicsatolási módokra, hogy megkülönböztessék a 4G-hálózatban bevezetett funkciót az 5G új architektúrájától, melyben az ilyen megoldások az alapvető konfiguráció részei lehetnek (hiszen az UPF-funkciókból tetszőleges számú köthető egymás mögé).

Minden kicsatolási módszer megőrzi azt az alapvető szabályt, hogy az alkalmazások a GTP-alagút után az „IP point of presence” mögött futnak, és csak azután manipulálhatják a hálózati forgalmat, miután az adatcsomagok elhagyták a 3GPP által specifikált hálózati domaint. (Ez röviden azt jelenti, hogy a mobilszolgáltató felelős a csomagok továbbításáért a saját rendszerén belül és az alkalmazások csak a rendszerből kilépő csomagokat manipulálhatják közvetlenül.) A 2G/3G/4G/5G-hálózatban használt központi anchor továbbfejlesztésével jöttek létre az edge computingot támogató kicsatolási módok, amelyeket a *4. ábra* illusztrál.

Az adatforgalom kicsatolásának egyik első felhasználási esete a hálózati forgalom optimalizálása. Több tartalomszolgáltató és globális felhőszolgáltató is telepít és konfigurál tartalomelosztó hálózatokat, amelyek a gerinchálózati forgalmat optimalizálják, illetve ezzel párhuzamosan javítják bizonyos tartalmak minőségét. A megoldás alapja, hogy a kicsatolási pont után a garanciát nem adó internet helyett, a garanciát nyújtó bérelt vonalakon történik a tartalom forgalmazása.

Vannak erre specializálódott cégek és természetesen a felhőszolgáltatóknak is vannak ide sorolható szolgáltatásai, pl. a Google Cloud Networking Services. Ide tartoznak még a különböző cégek SD-WAN (Software Defined Wide Area Networks) szolgáltatásai, melyek elsődleges célja a vállalatok számára biztosított privát hálózat, ami az Ipar 4.0 kontextusában egyfajta edge computing megoldássá fejlődött. Az ipar esetében az adatbiztonsági, rendelkezésreállási és kis késleltetési követelmények miatt a kritikus alkalmazások forgalmára kizárólag a helyi kicsatolást használják, míg az egyéb, például a monitoring forgalmak központi, adott esetben egy vállalat székhelyén lévő adatközpontba irányítódnak.



4. ábra Kapcsolati modellek a 3GPP-szabvány alapján.

Látható, hogy az 5G- és 6G-hálózatok heterogén kicsatolási pontjai más és más alkalmazásforgalom számára optimálisak. Ez vezet el minket a következő témakörhöz, mely a megfelelő kicsatolási pont megtalálását irányozza elő.

4.2. A megfelelő edge alkalmazás-szerver kiválasztása

Az adatkicsatoláson kívül a 3GPP-szabvány ajánlásokat fogalmaz meg az ún. edge application discovery-re, azaz arra, hogyan fogja a felhasználói eszköz a számára megfelelő edge alkalmazást kiválasztani.

Az első itt tárgyalt ilyen ajánlást a 3GPP SA2 testülete, azaz a hálózati architekúrát definiáló csoportja adja. A TS 23.548-ban bevezetik az Edge Application Discovery eljárást, amely arra hivatott, hogy a felhasználói készülék (User Equipment, UE) felfedezze az applikáció(k) (Edge Application Server(s), EAS) IP-címeit, a hálózatba telepített helyi vagy központi DNS-szolgáltatók segítségével. Ennek azonban az az előkövetelménye, hogy az UE és a rajta futó alkalmazások DNS-beállításai a megfelelő, mobilszolgáltató által megadott DNS-ek legyenek.

Egy második lehetőség a globális felhőszolgáltatók által szolgáltatott DNS használata, ami sok esetben felülírja a mobil hálózatban lévőt. A Google Androidja például titkosított lekérdezéseken keresztül a Google DNS-szolgáltatását használja, amíg más cégek, mint például a CloudFlare, hasonló szolgáltatást nyújtanak a vállalati adatbiztonság támogatásának céljából. Azaz az eszközök egy titkosított hálózaton kommunikálnak, függetlenül a mobilhálózattól.

Egy harmadik, naiv hozzáállás a legmegfelelőbb edge alkalmazás kiválasztására az lehet, hogy az alkalmazási rétegben, maga az alkalmazásszerver elküldi a lehetséges IP-címek listáját és a kliensalkalmazás ICMP- (ping-) protokoll segítségével választja ki a megfelelőt. Egy ilyen megoldás előnye többek között az lehet, hogy az alkalmazáskliens tudatosan választ és újraválaszt szervert, lerövidítve vagy teljesen kiiktatva ezzel a mobilitásból és hívásátadásból adódó potenciális kiesést. Többek között ebből a megfontolásokból javasolja a 3GPP SA6 missziókritikus alkalmazások csoportja az applikációs rétegben implementált edge application server discovery funkciót.

Amint az megfigyelhető, a telekommunikációs szabványosítási testületek és az IT-cégek által javasolt megoldások sokszor egymást kizáróak, egymás piacait diszruptálhatják.

4.3. Az edge-felhő

Manapság sokszor összemosódik az edge computing és az edge cloud fogalma, nem véletlenül. A felhőalapú szolgálatatások elterjedésével a felhő megjelent nem csak a mobil- és vezeték nélküli hálózatok infrastruktúrájának elemeként, hanem a hálózat felett nyújtott szolgáltatások (mobilapplikációk) futtatási környezeteként is.

Amint azt a bevezetőben is láttuk, jelenleg többféle megoldás létezik az edge cloud-ra. Az ETSI MEC a virtualizációs réteg fölé egy applikációs platformot vízionál, amit a mobiloperátor szolgáltathat az alkalmazás számára. Ez hasonló funkciókkal rendelkezne (pl. alkalmazási piac), mint a globális felhőszolgáltatók által jelenleg is üzemeltetett központi felhőplatformok. Kérdés, hogy az 5G-hálózatokba integrált edge computing megoldások esetén a mobilszolgáltatók fognak-e felhőszolgáltatási platformot fejleszteni az alkalmazásoknak, vagy azt valamilyen globális felhőszolgáltatóval közösen fogják nyújtani. Mindkettőre van példa, sokszor egyazon operátor esetén is, attól függően, hogy milyen alkalmazásról beszélünk, illetve, hogy az mennyire integrálódik a hálózattal.

Általánosan elfogadott, hogy a missziókritikus kommunikációt, nagy rendelkezésreállást és sokszor alacsony késleltetést igénylő alkalmazások esetén sokkal nagyobb mértékű lesz az integráció. Ennek egyik oka, hogy a szolgáltatás- és adatbiztonsági követelmények teljesítése érdekében sokszor a rendszer egészét kell tesztelni. A másik pedig, hogy a rendszerek szorosabb együttműködése szükséges a magas rendelkezésreállást biztosító architektúra és protokollok implementálásához.

Az Ipar 4.0 megoldások esetén a virtualizáció első lépéseként az ipari robotok logikáját egy külső számítási egységre helyezik át, ezzel elősegítve több robot együttes vezérlését és hatékonyabb koordinálását. E platform futtatja az ún. virtual Programmable Logic Controllert. Egy lehetséges következő lépés az ilyen célhardware-en futó vPLC-alkalmazást egy speciális felhőkörnyezetbe helyezni. Több ipari alkalmazás számára nem szükséges a speciális futtatási környezet, de szükséges az ipari biztonság, protokollok és szabványok támogatása. Ennek érdekében többen is saját ipari platformot fejlesztenek (pl. GE Predix, Siemens Industrial Edge), melyeket egy adott telephelyre, gyártósorhoz kihelyezve eleget tehetnek a késleltetési, biztonsági és rendelkezésreállási követelményeknek.

Érdemes tehát megfigyelni, hogy az edge computing környezetben is megtalálható a felhőalapú számítás rétegződése (IaaS, Container-aaS, Platform-aaS és Software-aaS), illetve, hogy ebben a rétegződésben minél feljebb megyünk, annál applikáció- specifikusabb platformokkal találkozunk.

4.4. Az edge-felhő menedzsmentje és orkesztrációja

Bár az ETSI MEC definíciója alapján az edge-felhő szolgáltatás szorosan integrálva van a hálózatba, az architektúra elmei közé tartozik a központi menedzsment és orkesztráció, mégis csak a telekommunikációs alkalmazások számára elkészült ESTI NFV MANO architektúrával találkozhatunk a gyakorlatban. Általánosságban elmondható, hogy sok vertikum ipari szereplői saját fejlesztésű komponenseket használnak az általuk fejlesztett alkalmazások menedzsmentjére. A feladat az infrastruktúra menedzsmentjének megoldása, illetve a kettő összehangolása marad.

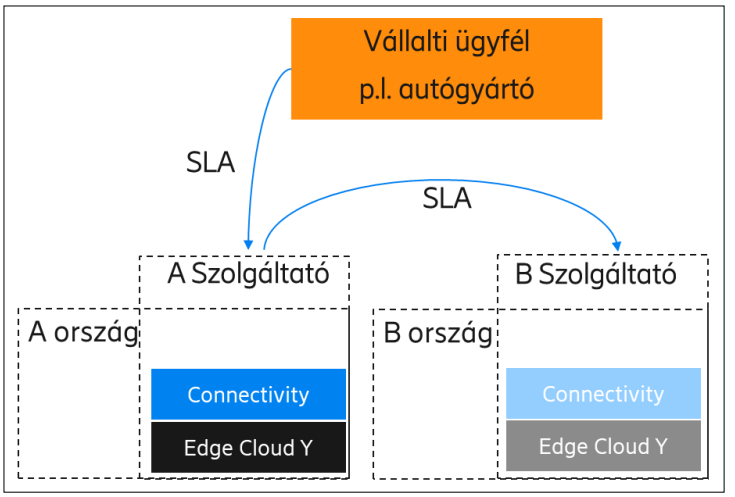
Az ETSI MEC, hasonlóan mint az alkalmazás esetén, az ETSI NFV alapján tartalmaz ajánlásokat az elosztott felhő-infrastruktúra menedzsmentjére, de mindezeket

megoldják a globális felhőszolgáltatók is. Az AWS esetében a Greengrass például probléma nélkül menedzselhető a központi felületről, csakúgy mint az Outpost. Hasonló egyedi megoldásokkal találkozunk az MS IoT Egde és a Google Anthos esetében. Szóval minden alkalmazásfajta más-más felügyeleti és telepítési rendszert használ.

Ezek a problémák nem csak a telekommunikációs iparágban merülnek fel. A probléma megoldását szolgáló technológiákat összefoglaló néven multi-cloud technológiáknak nevezzük.

A cégek nagy többsége több felhőtechnológiát is használ egyszerre. Példa: egy vállalat Gmail-t használ levelezésre, Microsoft Teams-et videókonferenciára és AWS Kubernetes clustert az általa fejlesztett alkalmazás futtatókörnyezetére. Ilyen esetekben legtöbbször elég, ha a felhasználó fennakadás nélkül használhatja ugyanazon felhasználói azonosítóját a különböző platformokon. Ennél összetettebb multicloud-technológiák szükségesek akkor, ha az alkalmazás maga támogat többféle platformkomponenst és az alkalmazást mozgathatjuk, például AWS- és Google-felhő között.

Az IT-világ számos opensource-projektje, például a Terraform, OpenNebula mellett a mobilszolgáltatók is foglalkoznak a kérdés megoldásával. A GSMA Operator Platform Group egy olyan, telekommunikációs szolgáltatók föderációján alapuló megoldást szorgalmaz, ahol a jelenlegi roaming-rendszerekhez hasonlóan az alkalmazásfuttatási-környezeti és adatforgalmi igénye megosztható a partnerszolgáltatóval, ezáltal biztosítva a megfelelő szolgáltatásminőséget.



5. ábra Egységes edge-szolgáltatás országok között.

4.5. Az edge computing, mint hálózati szolgáltatás és az edge exposure

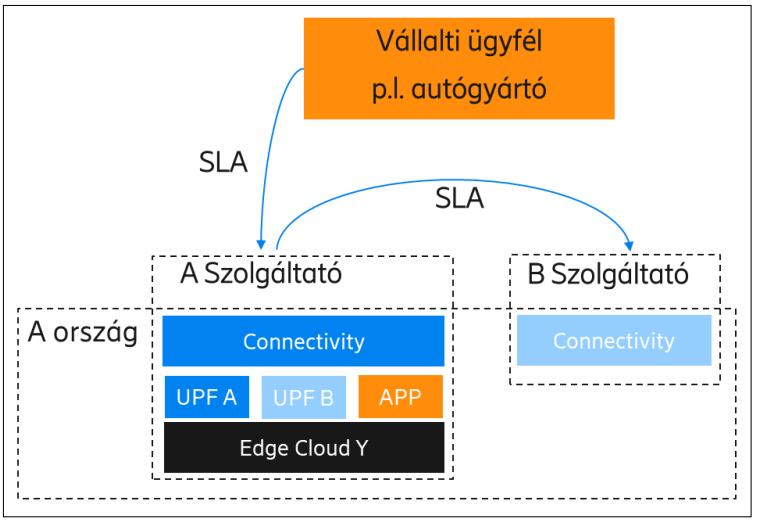
Amennyiben autóipari edge computing szolgáltatásról beszélünk, úgy felmerül az a követelmény, hogy az edge alkalmazás hálózati és felhőszolgáltatótól függetlenül elérhető legyen, ráadásul azonos minőségben, világszerte. Az ETSI MEC és a GSMA OPG, csakúgy mint sok másik vertikális ipari fórum, egy központosított feladatnak tekinti az alkalmazások telepítését. Ezek együtt kell tudjanak működni a globális felhőszolgáltatók már létező rendszereivel.

Ami az edge computingot különlegessé teszi, az, hogy az edge alkalmazások alapvető kommunikációs követelményként jelölik meg a kommunikáció minőségének biztosítását. A szolgáltatást alapvetően az alkalmazás készítője, forgalmazója nyújtja az alkalmazás felhasználóinak, mobilhálózatok felett, így a mobilhálózat által nyújtott kommunikációs kapcsolat egy szolgáltatásnak tekinthető.

Két kihívást tárgyalunk röviden ezzel kapcsolatban. Az első kihívás *(5. ábra)* a roaming esete, amikor az A országban lévő mobilszolgáltató ügyfele egy B országban roaming partneren keresztül csatlakozik. Az alkalmazás szempontjából ugyanazok a hálózati követelmények érvényesek mindkét esetben. Ugyanakkor a két hálózat esetében megoldandó, hogy az A hálózat értesítse B hálózatot a szolgáltatásminőségi követelményeiről, B hálózat pedig teljesítse azokat. Ez magában foglalja annak igényét, hogy az edge alkalmazást az A szolgáltató által üzemeltetett edge-felhőből a B szolgáltató által üzemeltett edge-felhőbe migrálják. Kérdés, hogy az alkalmazásfejlesztő számára mennyire lehet ezt láthatatlanná tenni, azaz, hogyan fog ez a legkevesebb adaptációt igényelni a fejlesztők részéről.

A másik kihívás *(6. ábra)* az országon belüli infrastruktúrára vonatkozik. Példaképpen egy repülőtér számára szeretne egy kiterjesztett valóság-szolgáltatást nyújtani egy startup. Ez esetben a vállalati ügyfél egy adott területen szeretne mobilszolgáltató-független edge computing-megoldást nyújtani. Ebben a sem üzleti, sem szabályozási szempontból nem kiforrott felállásban a technikai kihívást az jelenti, hogy lehet-e egy edge-felhővel (illetve annak menedzsment- és orkesztrációs rendszereivel) kiszolgálni a vállalati ügyfelet, több mobilszolgáltatón keresztül.

6. ábra Egy edge-felhővel kiszolgált alkalmazás, országon belül.



Példa: az egyik szolgáltató szerződésben áll a repülőtérrel, edge computing-szolgáltatást nyújt egy felhőszolgáltató bevonásával. A többi mobilszolgáltató ügyfelei számára pedig lehetővé teszi az edge forgalom helyi kicsatolását és átirányítását a megfelelő felhőrendszerbe. Ennek érdekében egy másik szolgáltató UPF-(speciális telekommunikációs IP-útvonalválasztó) funkcióját is telepítenie kell, ami nem bevett gyakorlat.

Mindkét megoldásban megjelenik az edge computing és a kapcsolat, mint szolgáltatás. A alapvető követelmények (alacsony késleltetés és sávszélesség) mellett megjelennek új típusú követelmények, például hálózatszelethez, helyhez, felhasználói típushoz (pl. autóba telepített egységek) kapcsolódóak, és mindenképpen specifikálni kell az alkalmazás számítási és memóriaigényét a felhőben. A mobil- és egyben edge computing-szolgáltatók közötti protokollrendszer követelményeinek kidolgozása jelenleg is zajlik a GSMA OPG-ben.

A potenciális alkalmazásokat és alkalmazási területeket megismerve látszik, hogy a legtöbb esetben szükség lesz arra, hogy az alkalmazásfejlesztő számára konfigurálható és ellenőrizhető legyen, hogy melyik alkalmazáshoz melyik felhasználói eszközök férhetnek hozzá, illetve, hogy adminisztrálja a hozzáférés mikéntjét és minőségét. Erre vonatkozóan találhatunk ajánlásokat a 3GPP SA6 munkacsoportjától és – többek között – az 5G ACIA ipari összefogás is ad ajánlást [7].

Míg az eddigi követelmények leginkább a management-aspektusra vonatkoznak, több helyen vizsgálják az edge szolgáltatások valós idejű konfigurálásának lehetőségét, például a szolgáltatásminőség dinamikus és igény szerinti beállítását a Network Exposure Function-on keresztül. (A Network Exposure Function lehetőséget ad harmadik fél által fejlesztett alkalmazások számára, hogy a hálózattal, API-k használatával interakciókat hajtsanak végre. A lehetséges további ilyen szolgáltatásokat az Edge Native Applications c. cikkben tárgyaltuk.)

5. Összefoglalás

A cikkben bemutattuk, hogy az edge computing egyfajta üzleti disztrupció a telekommunikációs és felhőszolgáltatás világában egyaránt. A két terület konvergenciájaként létrejövő edge computing területe több kihívást is támaszt mindkét oldal számára. A hálózatközeli forgalomkicsatolás problémája esetében láthattunk többféle tervezetet szabványos megoldására. A megfelelő edge alkalmazás megtalálásáért felelős funkcionalitás esetében láthattuk, hogy mind a hálózati, mind a globális felhőszolgáltatók és az internet világából is vannak megoldási javaslatok. Ezek legintenzívebb találkozási pontja a felhő-infrastruktúra és azon belül is annak menedzsment- és orkesztrációs rendszerei, illetve az itt felmerülő kérdéseire kifejlesztett, sokszor közös válaszok.

Az edge computing mint szolgáltatás az a terület, amely minden iparági szerelő számára nyújt lehetőségeket és kihívásokat. A felvázolt problémákból is látható, hogy sok megválaszolatlan kérdés van még és az is, hogy ezek a kérdések nem csak technikai természetűek, hanem üzleti és szabályozási területeket is érintenek. Bízunk benne, hogy a területen sok érdekes technikai innováció születik majd az elkövetkezendő években.

Hivatkozások

[1] Ericsson Mobility report, 2021. november:
https://www.ericsson.com/4ad7e9/assets/local/reports-papers/mobility-report/documents/2021/ericsson-mobility-report-november-2021.pdf

[2] Kovács B., Szilágyi L., Gera Z., Fábián G., Ferrari C.F.: Mesterséges intelligencia felhasználási esetek 5G hálózatokban. In: Híradástechnika HTE Infokom különszám, (2019), pp.2–7.

[3] Kovács B., Suskovics P., Terrill S., Wörndle P.: Creating the next-generation edge-cloud ecosystem. In: Ericsson Technology Review, 18 Februar (2020).

[4] Walsh, D., Walsh R.: AT&T to run its mobility network on Microsoft's Azure for Operators cloud, delivering cost-efficient 5G services at scale, Published by AT&T, 30 June 2021.
https://about.att.com/story/2021/att_microsoft_azure.html

[5] Hardesty L.: Telefónica Germany uses AWS, Ericsson to virtualize its 5G core, Fierce Wireless, 2 September 2020.

[6] 3GPP TS 23.548, Section 4.3 Connectivity Models.

[7] Exposure of 5G Capabilities for Connected Industries and Automation Applications. by ZVEI – German Electrical and Electronic Manufacturers' Association 5G Alliance for Connected Industries and Automation (5G-ACIA), a Working Party of ZVEI, Februar (2021).

A szerzőről

KOVÁCS BENEDEK edge computing szakértőként dolgozik az Ericsson Magyarországnál. Főbb felelősségi körei a mobilhálózatok és azon belül is az edge computing evolúcióját segítő mérnöki tevékenységek, prototípusok és tanulmányok elkészítése, a magyarországi leányvállalat kutatás és fejlesztési igazgatóságán a technológiai és innovációs csoport vezetése, valamint felelős az egyetemekkel és akadémiai intézményekkel való kapcsolattartásért. MSc diplomáját mérnökinformatikusként, PhD fokozatát matematikus mérnökként szerezte meg a Budapesti Műszaki és Gazdaságtudományi Egyetemen.

Mit ad az átlagfogyasztónak az 5G?

BARTOLITS ISTVÁN
Nemzeti Média- és Hírközlési Hatóság
bartolits@nmhh.hu

Kulcsszavak: 5G-hálózatok, 5G NSA- és SA-rendszer, átlagos fogyasztó, új 5G-szolgáltatások

A cikk arra keresi a választ, hogy a vertikumok számára sok előnyt ígérő 5G-hálózatok mit nyújtanak most és a jövőben az átlagos fogyasztók számára. A cikk első része az 5G-rendszerek jelenlegi piaci helyzetét mutatja be, majd kitér a piaci bevezetés nehézségére és az itt felmerült „tyúk vagy a tojás" probléma megoldására. A következő részben bemutatja azokat az alkalmazásokat, amiket már az 5G non standalone rendszer alatt igénybe lehet venni, majd kitér a távolabbi jövő, az 5G standalone rendszer sokoldalú szolgáltatásaira.

1. Bevezetés

Az ötödik generációs mobilhálózatok szabványosítási munkái során sok hír jelent meg arra vonatkozóan, hogy mennyiben fog ez a rendszer többet tudni, mint a korábbi mobilrendszerek, beleértve a jelenleg domináns LTE-hálózatokat is. A világ 5G-lázban égett, egymás után indultak meg a frekvenciaárverések az 5G-rendszerek számára is használható rádióspektrumok vonatkozásában, és megkezdődött a hálózatok kiépítése. Közben egyre több tanulmány foglalkozott azzal, hogy az 5G-hálózatok alapot fognak adni más iparágak rendszereinek kiszolgálására, s egyre nyilvánvalóbbá vált, hogy ezen iparágak számára – amiket összefoglaló néven vertikumoknak neveznek – valóban nagy lehetőségeket fog nyújtani az 5G-rendszer.

Kiszolgálja majd az Ipar 4.0 igényeit, ráépíthető lesz az e-egészségügy, 5G-magánhálózatokkal látják majd el az olyan közlekedési csomópontokat, mint például a repülőterek, vasúti rakodók és kikötők, valamint a gyárakat, üzemeket és hasonló komplexumokat? Az is kiderül ezekből a tanulmányokból, hogy az említett alkalmazások fognak leginkább bevételt termelni a szolgáltatók számára. Arra a kérdésre azonban, hogy mit ad az utca emberének, az egyedi előfizetőknek – azaz az átlagfogyasztóknak – az ötödik generációs rendszer, ezek a tanulmányok nem adnak valódi támpontot. Éppen ezért ebben a cikkben erre a kérdésre keressük a választ.

A második szakaszban egy körképet adunk arról, hol is tart a cikk megírásának az időpontjában az 5G-rendszer elterjedése és a végberendezések hozzáférhetősége. Ezt követően az 5G-rendszer ígéreteit villantjuk fel, rámutatva arra az üzleti ellentmondásra, hogy a nagy beruházást igénylő váltást nem lehet a szabványosítás lezárásáig visszatartani, mert közben fel kellene futtatni a végberendezések piacát. Ennek a „tyúk vagy tojás" problémának a megoldását és a következményeit, korlátait mutatja be a negyedik és ötödik szakasz. A hatodik és a hetedik szakasz azt mutatja be, hogy a korlátok ellenére milyen új lehetőséget ad az 5G-rendszer már most is az átlagfogyasztók számára, az 5G teljes szabványosítása után pedig ezek a lehetőségek milyen jellegű alkalmazásokkal, szolgáltatásokkal fognak kiteljesedni.

2. Hol tart a világban az 5G?

Először tekintsük át, hogy 2022 januárjában hol tart az 5G elterjedése a világban a végberendezések, a kiépülő hálózatok és az előfizetők tekintetében.

A Global Mobile Supplier Association (GSA) 2022. januári jelentése szerint 180 gyártó jelentett be 5G-hálózatra kapcsolódni képes eszközöket, összesen 1257 különböző típust [1]. Ezek közül 857 típus kereskedelmi forgalomban is elérhető. A mobiltelefonok esetében 614 típus látott eddig napvilágot, ezek közül 545 típus szabadon megvásárolható, a többi 69 típus nem kereskedelmi forgalmazásra készült. 210 Fixed Wireless Access (FWA) eszköz is rendelkezésre áll, melyek közül 98 már nagykereskedelmi piacon is elérhető, ezek egy része beltéri, másik része pedig kültéri telepítésre is alkalmas. Elérhető már 81 típus ipari, illetve vállalati router, bridge és modem is, melyek 5G-képesek. Az akkumulátoros hotspotok között is már 55 típus található meg. 28 notebook és ugyancsak 28 tablet is megjelent a piacon 5G-interfésszel. 11 gépjárműbe telepíthető eszköz is elérhető a piacon, ezek routerek, modemek és hotspotok, valamint 8 USB-modem is megvásárolható már az 5G-képes eszközök között. Mindebből látszik, hogy – ha nem is olyan mértékben, mint a 4G-hálózatok tekintetében –, van már készülék- és eszközválaszték az 5G-re csatlakozó végberendezések terén. A felfutásuk pedig gyorsabbnak néz ki, mint annak idején a 4G-képes eszközöké.

Az 5G-hálózatok bevezetése is erőteljesen növekszik. 2022 elejére 72 országban 187 szolgáltató indított 5G-mobilszolgáltatást [2]. 83 szolgáltató 45 országban már a 3GPP-szabvány szerinti helyhez kötött vezeték

nélküli hozzáférést, az 5G FWA-szolgáltatást is bevezette. A GSA nyilvántartja azt is, hány szolgáltató kezdte meg a beruházásokat 5G-rendszer kiépítésébe valamilyen szinten (tesztelés, tervezés, hálózatfejlesztés szintjén). A már kereskedelmi forgalomban lévő hálózatokkal együtt ez már 487 szolgáltatót számlál összesen 145 országban.

Ezzel együtt természetesen az 5G-előfizetők száma is jelentősen növekszik. Erre vonatkozólag erősen eltérő adatok olvashatók a különböző jelentésekben, aminek több oka is van. A legfőbb ok talán az, hogy keverednek a statisztikai besorolások. Van, aki az eladott 5G-képes eszközök alapján adja meg az előfizetőszámot, bár az egyáltalán nem biztos, hogy az eszköz valaha is felcsatlakozott volna 5G-hálózatra. Egy másik fontos tényező, hogy 5G-képes eszközről vagy 5G-képes okostelefonról beszélünk, hiszen éppen az 5G esetében ezek mennyisége már szignifikánsan eltérhet egymástól. Ha a szolgáltató ezekben a statisztikai besorolásokban egyértelműen állást foglal, akkor is kérdéses, hogy milyen forgalmi adatok esetén lehet azt mondani, hogy az előfizető 5G-hálózatot használ. Már akkor is, ha egyszer felcsatlakozott az 5G-hálózatra, miközben szinte a teljes forgalma LTE-hálózaton bonyolódik, mert még nincs 5G-lefedettsége? Ezek mind olyan kérdések, amelyek megnehezítik az 5G-használók pontos számának megadását.

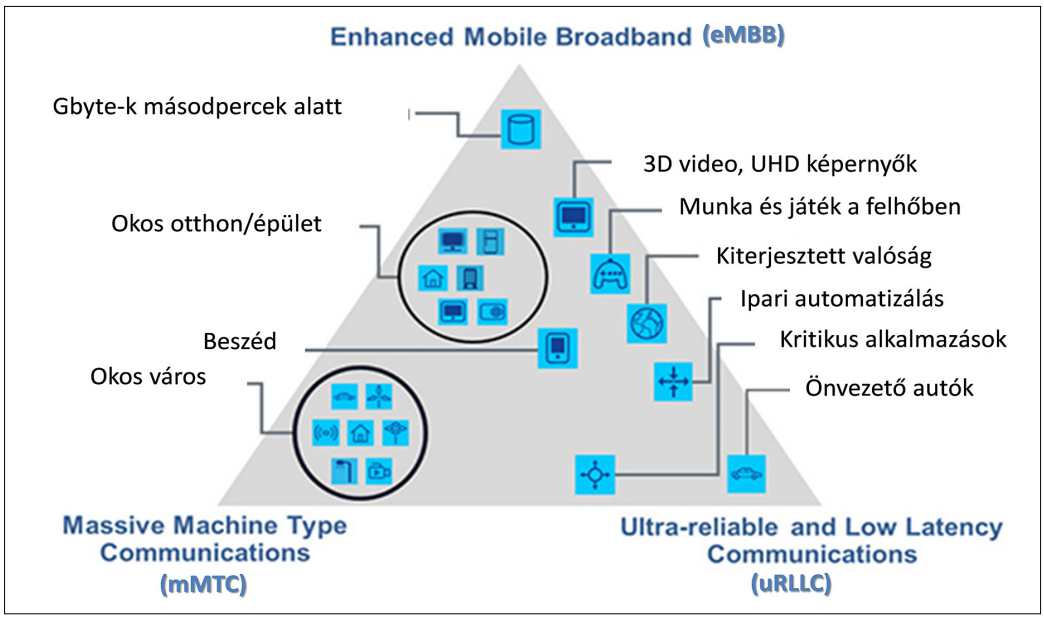
Ennek ismeretében talán már nem meglepő, hogy nagyon eltérnek a statisztikai adatok. A Statista adatai szerint [3] 2020 végén már 236 millió, 2021 végén pedig 554 millió 5G-előfizető volt a világon. Előrejelzésük szerint 2022 végére ez a szám már elérheti az egymilliárdot, 2025 végére pedig a 3 milliárdot. Hasonló nagyságrendet állapított meg az Ericsson Mobility Report 2021 novemberében megjelent kiadása [4], mely 2021 végére 660 millió előfizetőt jelzett, felemelve a 2021. júniusi kiadás 580 milliós előfizetői előrejelzését. A megnövelt előrejelzést a vártnál is erősebb kínai és USA-beli előfizetési hullámmal indokolták. A Global Mobile Supplier Association még nem tett közzé év végi adatot, de a 2021 harmadik negyedévének végén már 438 millió előfizetőt jelzett. Az eltérő adatok bizonytalansága ellenére is annyit azért ki lehet mondani, hogy a 2022-es év indulásakor mintegy félmilliárd 5G-előfizető lehetett a világon, ami azért meglehetősen komoly szám.

A következőkben tehát arra keressük a választ, hogy mit ad az 5G-hálózatok szolgáltatási köre jelenleg és a jövőben ennek a félmilliárd, illetve távlatilag a 3-5 milliárd 5G-előfizetőnek.

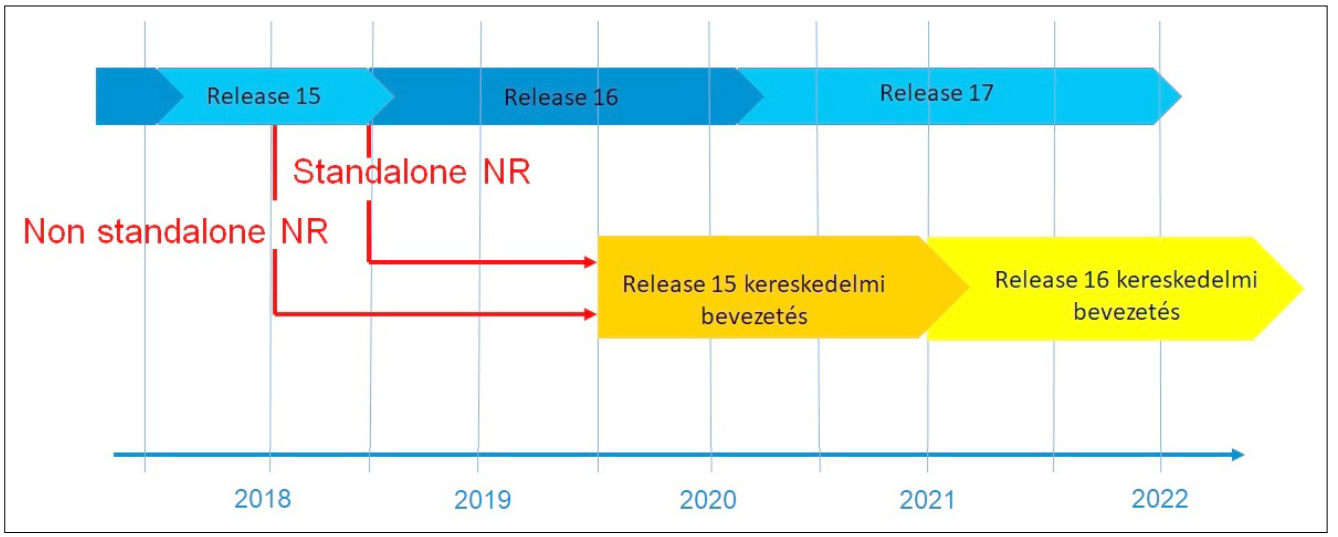
3. Az 5G ígérete és megvalósítása

Az ötödik generációs rendszerek specifikációja hosszú előzményekre tekint vissza, akárcsak a korábbi generációk előkészítése. Az ITU már 2012 körül megkezdte a vizsgálatát a 2020-as évek mobil rendszereivel szemben támasztandó követelményeknek az „IMT 2020 and beyond" program keretében. 2015-ben az ITU-T egy fókuszcsoportot hozott létre az IMT-2020 részletes specifikációjára (Focus Group on IMT-2020). Ez a fókuszcsoport – melynek munkájában a szerző is részt vett –, 2015 májusától 2016 decemberéig kilenc dokumentumban foglalta össze mindazokat az alapelveket, amiket az ötödik generációs hálózatok specifikációjánál figyelembe kell venni. Az IMT 2020 and beyond és a Focus Group on IMT-2020 munkájának az eredményeként már akkor nyilvánvalóvá vált, hogy az 5G-rendszerek minőségi változást hozhatnak a 4G-rendszerekhez képest, azonban a rendszer bevezetése nem lesz úgy megoldható, hogy egyből a teljes tudásával, az összes képességével induljon el rajta a szolgáltatás, mert ez hatalmas invesztíciót jelentene a szolgáltatóknak úgy, hogy a bevételi oldal még szinte nem hoz semmi eredményt.

Az IMT 2020 and beyond által kitűzött célokat ekkor foglalta az ITU az úgynevezett 5G-követelményháromszöget formáló ábrába, aminek a három csúcsában a három kiemelt követelmény állt *(1. ábra)*.



1. ábra
Az 5G-rendszerrel szemben támasztott követelményháromszög (Forrás: ITU-T 5G Focus Group).



2. ábra A 3GPP szabványosítási menetrendje 2018–2022. (Forrás: 3GPP)

Ezek: az igen nagy sávszélességű átvitel (eMBB – enhanced Mobile BroadBand), a tömeges IoT-eszközök kiszolgálása (MMTC – Massive Machine Type Communication) és a nagy megbízhatóságú, kis késleltetésű kommunikáció (uRLLC – Ultra-Reliable and Low Latency Communications).

Nyilvánvaló volt, hogy ezt a hármas követelményrendszert egy szolgáltatáson belül egyszerre nem lehet kielégíteni, de nem is ez volt ennek a követelményháromszögnek a célja. Ugyanakkor az egyértelmű volt, hogy az 5G-rendszer egészének alkalmasnak kell arra lennie, hogy ezeket az igényeket – ha nem is egy szolgáltatáson belül – egyszerre tudja kiszolgálni. Már a specifikációs munka ezen szakaszában tehát kirajzolódott, hogy az 5G-rendszer szabványosítása egy igen hosszú folyamat lesz, s csak ennek a folyamatnak a befejezése után jelennek meg mindazok az előnyök, amiket az 5G-rendszer végső változatának a kiépítése valóban el fog hozni.

Ezzel azonban egy nagy ellentmondás is keletkezett: ha az 5G-rendszer csak a szabványosítás lezárása után lesz teljes értékű, ezért addig el sem indítja a szolgáltatást a szolgáltató, akkor addig nem fog felfutni a végberendezések gyártása sem, nem lesznek készülékek a piacon, nem alakul ki a korai felhasználók széles tábora, akik már az 5G-rendszerek kiépítésének a korszakában belépnek az előfizetők táborába. Erre az ellentmondásra valamilyen kompromisszumos megoldást kellett találni.

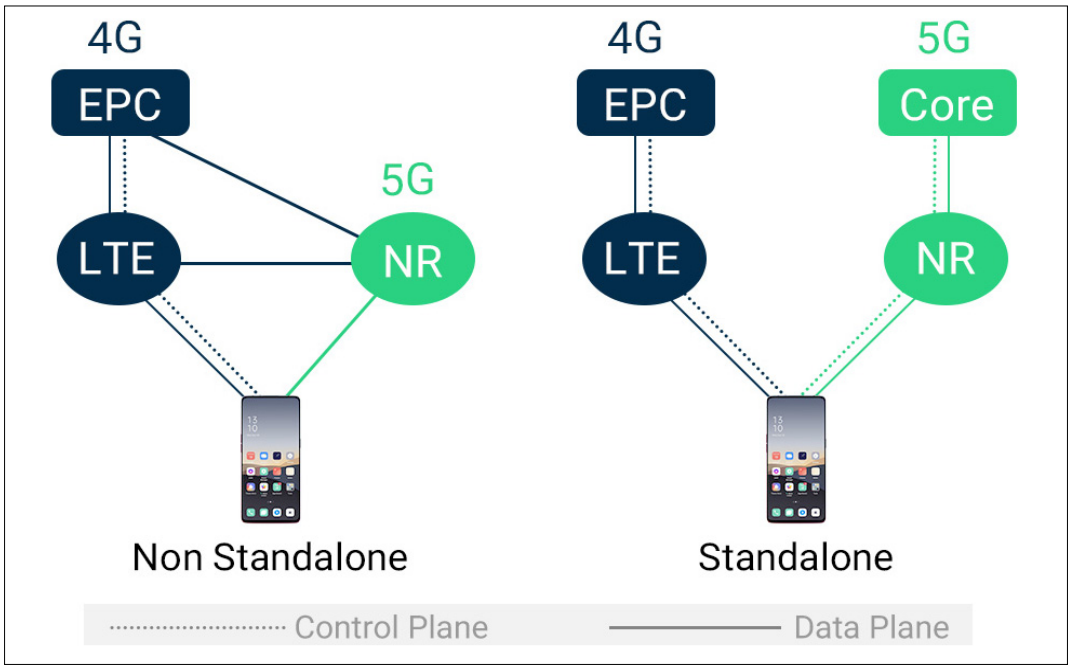
4. A „tyúk vagy tojás” probléma feloldása – az 5G Non Standalone rendszer

Ennek a „tyúk vagy tojás” problémának a feloldására született meg a 3GPP-ben (a mobil rendszerek szabványosítását összefogó szervezetben) a gondolat, hogy a három kiemelt követelmény közül ki kell emelni a nagy sávszélességű átvitelt, és az 5G rádiós interfészének a specifikálásakor előre kell hozni egy olyan megoldás kidolgozását, amely már az LTE-hálózatokra ráépíthető és korlátozottan ugyan, de 5G-képes szolgáltatást nyújt. Ezt a megoldást – mivel ekkor valójában nem önálló 5G-rendszerről, hanem egy 4G-hálózatra épített 5G-rádiórendszerről van szó – elnevezték *5G non-standalone (5G NSA)* rendszernek. A gondolat az volt mögötte, hogy így már el lehet indítani a nagyobb sávszélességet kezelni képes 5G-rádiórendszert, az 5G NR-t (5G New Radio) az LTE-hálózatokra telepítve. Ez megadja a nagyobb sávszélességű forgalmazás élményét azoknak az előfizetőknek, akik ezért hajlandóak 5G-képes okostelefont vásárolni, és ugyanakkor ennek az 5G NSA-rendszernek a kiépítése jóval hamarabb megvalósítható és jelentősen kisebb befektetést igényel. Az is igaz persze, hogy ezzel az előfizetők még nem azt az 5G-rendszert kapják a kezükbe, amiről a korábbi ígéretek szóltak.

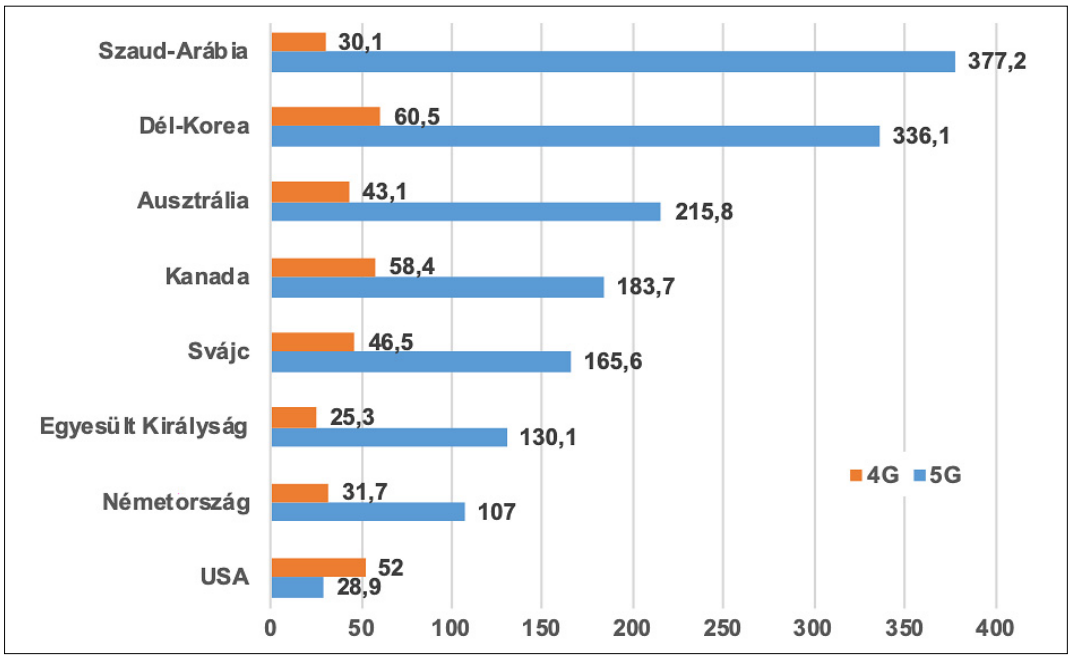
A non-standalone rendszer specifikálása után fél évvel elkészült a 3GPP-ben az *5G standalone (5G SA)* rendszer rádiós interfésze is, mellyel le is zárult az 5G-rendszer első, Release 15 névre hallgató fázisa *(2. ábra)*.

A két rendszer közötti különbséget sematikusan ábrázolva a következő oldali, *3. ábrán* láthatjuk. Az 5G NSA rendszerben a negyedik generációs EPC (Evolved Packet Core) a maghálózat továbbra is, a teljes vezérlés ezen keresztül történik. Az 5G-képes készülék is az LTE-hálózattal kommunikál, csak az adatok forgalmazása történik az 5G NR-rendszeren keresztül.

Az 5G SA-rendszerben már kiépítésre kerül az 5G-maghálózat és annak a komplex vezérlési síkja is. Itt az 5G NR mind a vezérlési sík, mind az adatsík szempontjából csak az 5G-maghálózathoz kapcsolódik. Az okostelefon fel tud kapcsolódni a 4G- és az 5G-hálózatra egyaránt, amennyiben mindkettő rendelkezésre áll. A két hálózaton viszont eltérő képességekkel, eltérő szolgáltatási körrel találkozik, itt már érvényesülnek az 5G-rendszer ígéretei is abban az esetben, ha az 5G-maghálózat tartalmazza az összes ehhez szükséges funkciót, aminek egy része már a Release 16 szoftver verzióban is megjelenik, de lényegében a Release 17 csomagban és esetleg egy kis részben a Release 18 szoftver verzióban válik teljessé.



3. ábra
Az 5G NSA- és az 5G SA-rendszer sematikus felépítése.



4. ábra
A 4G-rendszer és az 5G-rendszer átlagos sebessége néhány országban Mbit/s-ban.

(Forrás: Statista, 2020. szeptember)

5. Az 5G NSA-rendszer korlátai

A fentiekből következik, hogy a világon szinte az összes nyilvános 5G-hálózat non standalone hálózatként indult el, s jelenleg is alig másfél tucat rendszer állt át az 5G standalone üzemmódra. Mindhárom hazai 5G-szolgáltatónk is 5G NSA-hálózatot üzemeltet, még ha ezzel a hírrel így nem is találkozunk a reklámokban. Ez teljesen logikus lépés, fokozatosan lehet utána kiépíteni az 5G-maghálózatot és átállítani az 5G NSA-rendszert az 5G SA-működésmódra.

Mi viszont most azt vizsgáljuk, hogy mit ad az átlagfogyasztónak az 5G-rendszer, és az eddigiekből jól kiolvasható, hogy az a legalább félmilliárd előfizető, aki 5G-szolgáltatásokat vesz igénybe, többségében 5G NSA-rendszeren kapja a szolgáltatást.

Ez azonban behatárolja a lehetőségeket is. Az 5G NR ugyan nagyobb sávszélességet nyújt, mint az LTE-rendszer rádiórendszere, de ezt a forgalmat utána a 4G-maghálózat továbbítja. Amíg annak a kapacitása alkalmas ennek az egyre növekvő sávszélesség-igénynek a kezelésére, addig a sebesség szempontjából az előfizetői élmény valóban pozitív. Amikor azonban már sok 5G-képes végberendezés használja ki az 5G NR nagyobb sávszélességét, akkor a forgalmat már a maghálózat kapacitása kezdi korlátozni. Ez sok szolgáltató esetében nem okoz problémát, de a megavárosokban, ahol a lélekszám 10 millió feletti – s ma már több mint 40 ilyen város van a világon – már a 4G-forgalom is kitölti az LTE-hálózatok kapacitását, itt bizony égetően szükséges az 5G SA-rendszer megvalósítása. Nem véletlen, hogy már mindhárom nagy kínai szolgáltató (a China Telekom, a China Mobile

és a China Unicom) már 5G SA-rendszert üzemeltet, hiszen a megavárosok közül nyolc található Kínában. Az USA-ban is hamar átállt a feltörekvő T-Mobile US erre a rendszerre, de így működik a Telstra hálózata is Ausztráliában és a KT Dél-Koreában, valamint még vagy egy tucat kisebb szolgáltató.

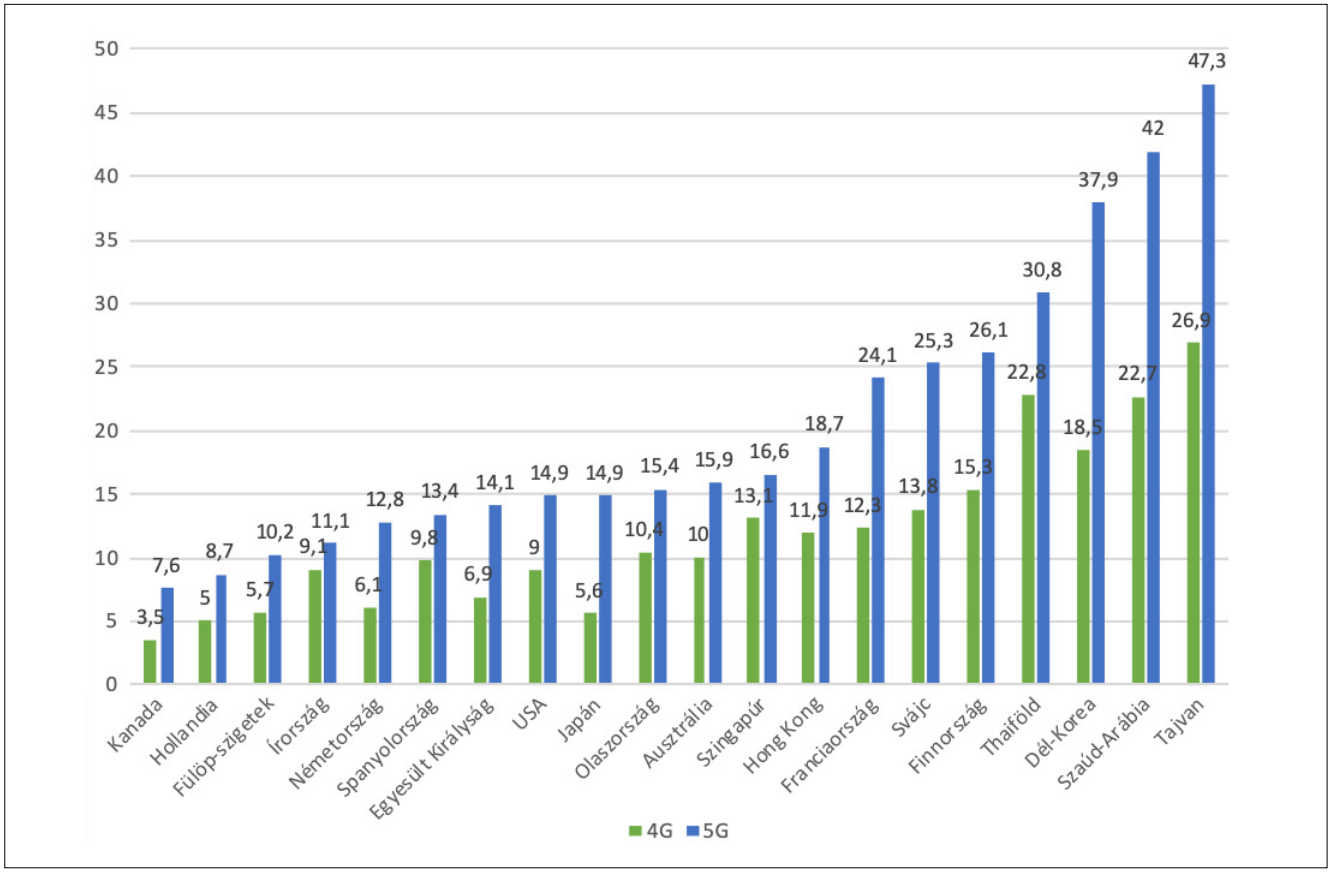
A gyakorlati adatok a sávszélesség növekedésére vonatkozó élményt viszont – legalábbis átlagosan – visszaigazolják. Ez már az első rendszerek elindítása után érzékelhető volt. A Statista 2020. szeptemberi adatai alapján [5] például Dél-Koreában a 4G átlagos 60,5 Mbit/s-os sebessége helyett az 5G-hálózaton 361,1 Mbit/s-os átlagot mértek, Svájcban 46,5 Mbit/s helyett pedig 165,6 Mbit/s-ot *(4. ábra)*. Az OpenSignal a havi adatforgalmazást vizsgálta [6] a húsz vezető 5G-országban, és azt találta, hogy az 5G-előfizetők havonta 1,5–2,7-szer nagyobb adatforgalmat bonyolítanak, mint a 4G-előfizetők *(5. ábra)*.

6. Jellemző sávszélesség-igényes alkalmazások

Az előfizetők egy része számára – akik számára nem fontos a nagy sávszélesség – az 5G NR nagyobb sebessége csak gyorsabb válaszidőt, esetleg jobb minőséget jelent. Ugyanakkor sokan kimondottan arra vártak, hogy nagyobb sávszélességű hálózaton tudjanak mobil adatforgalmat bonyolítani. Az ő számukra ugyanis már az 5G NSA-rendszer is komoly segítséget jelent. Melyek azok a sávszélesség-éhes alkalmazások, amiket az 5G-képes okostelefonnal rendelkező előfizetők használnak? Nos, erre vonatkozóan is készültek statisztikák, most ezek hivatkozása nélkül inkább csak a leginkább jellemzőket emeljük ki.

Sokan használják már a nagyobb felbontású (4K, 8K) videoátvitelt, s terjed a 360 fokos videofelvételek nézettsége is. Egyre növekszik az élő sportesemények mobiltelefonon történő követése, és itt már megjelenik a nagy felbontás és a választható nézet képessége is. Ugyancsak egyre növekvő tábora van a streamelt játékoknak és a felhőnatív játékoknak, amik használata már a függőség határát súrolja, van akinél már át is lépi azt. Az 5G-lefedettség növekedésével egyre inkább megjelennek a „connected car” szórakoztató rendszerek, melyre egy egész iparág szakosodhat a közeljövőben. Ezzel rokon – bár nyilván szerényebb igényekkel – a zenehallgatás mobil készüléken át bárhol, bármikor élménye. Terjed a kiterjesztett valóság (AR) használata és a virtuális valóság (VR) élménye is, aminek az igénybevétele igencsak sávszélesség-éhes. Több országban bevezették a turisztikai nevezetességek 5G-alapú információs rendszerének működtetését, ezzel téve teljessé a meglátogatott helyszínekről szóló tájékoztatást. Sorra születnek a felhasználói szintű IoT-alkalmazások is, amik ugyan szerény sávszélességet vesznek igénybe, de a nagy számosságuk miatt mégis tényezővé válhatnak.

5. ábra
A 4G- és az 5G-rendszer előfizetőinek átlagos havi adatfogyasztása Mbyte-ban.
(Forrás: Opensignal, 2021. január-márciusi átlagok)



Az alkalmazások közül kiemelkedik egy nagykereskedelmi szolgáltatás, melyet tipikusan 5G-rendszeren nyújtanak a szolgáltatók: a Fixed Wireless Access. Az év végén már 65 szolgáltató nyújtott ilyen szolgáltatást 35 országban és ez a szám egyre bővül. Elsősorban a külvárosi és rurál területek ellátását lehet vele jól, hatékonyan megoldani, persze versenyeznie kell az FTTP- és FTTH-megoldásokkal, ahol ezek nagyon elterjedtek, ott inkább csak kisegítő jelleggel használják.

7. Szolgáltatások az 5G SA-hálózatokon

Az 5G-hálózatok a fentiek alapján az igazi potenciáljukat a szabványosítás következő szakaszában, a Release 17, de még inkább a Release 18 verzió telepítése után fogják elérni. Ekkor már rendelkezésre fog állni a követelményháromszög teljes rendszere, és az 5G-hálózatokban a hálózatszeletelés (Network Slicing) bevezetésével kiteljesedhet az 5G-szolgáltatások köre. A legtöbb tanulmány ugyanazt emeli ki, miszerint ezeknek az új szolgáltatásoknak elsősorban az ipari alkalmazások körében lesz létjogosultsága, de már most látszanak olyan fejlesztések, amelyek az átlagfogyasztó számára is hoznak új lehetőségeket.

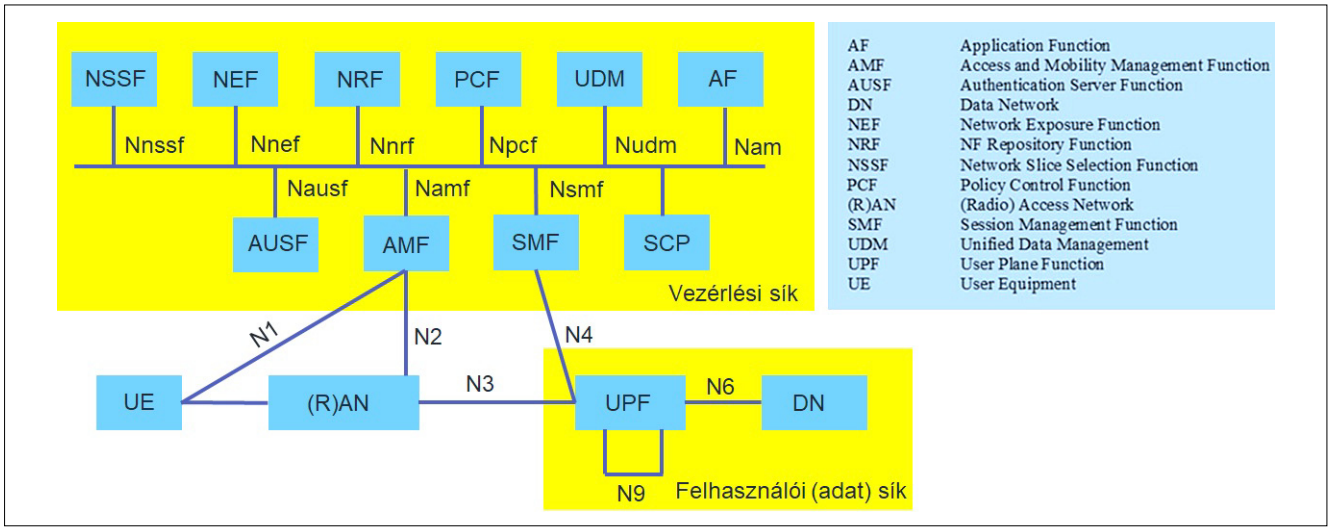
Ez derül ki azokból a jelentésekből is, amelyek a felhasználói piacra készülő fejlesztéseket jelzik előre [7,8]. Általában több csoportba sorolják be a fejlesztés alatt lévő alkalmazásokat, így jellemzően az autózás, a szórakoztatás, az okos otthon és IoT, a játékok és az AR/VR, a vásárlás és az immerzív kommunikáció, valamint a szélessávú alkalmazások csoportokba. Mindegyikben vannak már ígéretes eredmények, ilyen például a gépjárművek kiterjesztett valósággal kiegészített szélvédője, ahol a valós látvány mellett különböztető figyelmeztető feliratok, tájékoztató információk jelennek meg. Dolgoznak a valós idejű fordítórendszeren is, ahol a két eltérő nyelvű beszélgető között automatikus fordítást végez az 5G-rendszerre épített alkalmazás. Az 5G-vezérelt drónszállítás is a fejlesztendő alkalmazások között van, de már foglalkoznak a háromdimenziós hologramos hívás megvalósításával is, ami persze távlati terv, és az is lehet, hogy csak a 6G világában lesz kereskedelmi alkalmazás belőle, de már az 5G platformján is próbálkoznak a fejlesztésével.

A fenti példák jól mutatják, hogy lesznek olyan alkalmazások és szolgáltatások, melyek az átlagfelhasználók számára is hasznosak lesznek. A távolabbi jövőben, a vertikális ágazatok kiszolgálásának kiteljesedésekor pedig természetes módon fognak új lehetőségek nyílni az átlagfogyasztók számára is. Ilyenek lesznek például a B2C- és a B2B2C-alkalmazások, ezek végén ugyanis a fogyasztó áll és számára (is) előnyös lesz ezek igénybe vétele. A virtuális vásárlás élményét a haptikus vásárlási élmény bevezetésével lehet még növelni az 5G-szolgáltatásokon keresztül. Így a látvány és a hang mellett már a kinézett termék tapintása is a vevő rendelkezésére áll majd. Hatalmas potenciál rejlik az e-egészségügyi szolgáltatások kifejlesztésében, ami tipikusan mutatja, hogy a kiszolgált vertikális piac mi módon csatlakozik az előfizetői szegmenshez. Az átlagfelhasználó számára sok lehetőséget rejt az intelligens közlekedés területe is, ahol számtalan előnyös alkalmazás segíthet az autós, az önvezető járműves és a tömegközlekedés terén egyaránt.

Joggal merül fel a kérdés, hogy ha ezeket az alkalmazásokat utólag, egyesével fejlesztik ki az 5G-rendszerekhez, akkor mekkora nehézséget jelent majd a már működő hálózatokban történő implementációjuk. Erre már az 5G-rendszer referenciamodelljének a kidolgozáskor gondoltak. Mint ismeretes, az 5G-hálózatban alapelvként jelenik meg a vezérlési sík (control plane) és az adatsík elválasztása *(6. ábra)*.

A vezérlési sík referenciamodelljének kialakításakor már arról is gondoskodtak, hogy a natív felhőimplementálást segítsék, ezért a vezérlési sík egyes blokkjai mind önálló funkcióval rendelkeznek és nincsenek benne összevont, vegyes rendszerelemek. Ezen funkcionális blokkok egyike, az Application Function (AF) kimondottan az 5G-alkalmazások kezeléséért felelős, ennek a modulnak

6. ábra Az 5G hálózat referenciaábrája. (Forrás: ETSI TS 123.501.)



a segítségével lehet majd az egyes alkalmazásokat az 5G-hálózati rendszerbe telepíteni és működtetni. Mivel az 5G-rendszer kialakításakor a gyors alkalmazás-implementáció a kezdetektől kiemelt szempont volt, így a teljes szabványosítási folyamat során ezt figyelembe vették az ITU, az ETSI és a 3GPP szakemberei [9].

8. Összefoglalás

A cikkben megvizsgáltuk, hogy az 5G-hálózatok jelenlegi fejlődési szintjén milyen lehetőségekkel találkozik az átlagfogyasztó. Az 5G szabványosításának ütemtervéből jól látható, hogy a jelenlegi hálózatok többsége még nem önálló 5G-rendszer, hanem csak a 4G-hálózathoz illesztett 5G NR-rendszerelemek összessége. Ez is többletet nyújt az előfizetőknek, ami az adatátviteli sebességet illeti, de még nem rendelkezik az 5G-rendszer összes tervezett előnyével. Az előfizetők számára az igazi alkalmazási lehetőségek az 5G SA-rendszerek kialakítása után jelenhetnek meg, ez csak 2-4 év múlva várható. Az 5G-rendszer és a ráépülő alkalmazások akkor fognak az előfizetők számára is teljesen új, ma még talán meg sem jósolható szolgáltatásokat nyújtani.

Hivatkozások

[1] 5G Ecosystem Report; January 2022. Global mobile Suppliers Association, 2022. Elérhető: *https://gsacom.com/technology/5g/*

[2] 5G Market Update; End December 2021. Global mobile Suppliers Association, 2022. Elérhető: *https://gsacom.com/technology/5g/*

[3] Forecast number of mobile 5G subscriptions worldwide from 2019 to 2025, Statista. Elérhető: *https://www.statista.com/statistics/760275/5g-mobile-subscriptions-worldwide/*

[4] Ericsson Mobility Report, November 2021; Ericsson. Elérhető: *https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2021*

[5] How fast is 5G?, Statista, October, 2020. Elérhető: *https://lb-aps-frontend.statista.com/chart/22723/average-5g-and-4g-download-speeds-in-selected-countrie/*

[6] 4G.DE; Opensignal: Deutsche Nutzer verbrauchen 6,3 GB monatlich. Elérhető: *https://www.4g.de/news/opensignal-deutsche-nutzer-gb-12280/*

[7] 5G consumer potential – Busting the myths aroung the value of 5G for consumers; Ericsson Consumer & IndustryLab Insight Report, May 2019. Elérhető: *https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/5g-consumer-potential*

[8] Five ways to a better 5G – Key trends influencing consumer adaption of 5G; Ericsson Consumerlab, May 2021. Elérhető: *https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/five-ways-to-a-better-5g*

[9] 5G; System architecture for the 5G system (5GS), ETSI TS 123 501 Techical Specification, Version 16.11.0, Release 16; January 2022.

A szerzőről

BARTOLITS ISTVÁN 1978-ban szerzett villamosmérnöki, 1980-ban híradástechnikai szakmérnöki diplomát, 1983-ban pedig egyetemi doktori fokozatot a BME Villamosmérnöki Karán. 20 éven keresztül a BHG Fejlesztési Intézet fejlesztőmérnöke, fejlesztési osztályvezetője, majd projektmenedzsere volt a távközlés területén. Emellett 1993–1999 között a hírközlésért felelős miniszter tanácsadó testületének, a Távközlési Mérnöki Minősítő Bizottságnak az alelnöke volt. 1998 óta dolgozik a Nemzeti Média- és Hírközlési Hatóságnál, illetve jogelődeinél. Először elnökhelyettesi tanácsadó, majd osztályvezető volt. 2010 óta a Technológiaelemző Főosztály vezetője. A szabályozási munka támogatása mellett tevékenységi körébe tartozik az új technológiák, szolgáltatások megismerése, elemzése és az általuk felmerülő szabályozási kérdések azonosítása. Több nemzetközi szervezetben (ITU-T SG 13, ITU-T Future Networks 2030 Focus Group, Broadband Forum, BEREC, EU CNECT Expert Group 112) az NMHH, ill. Magyarország szakértő képviselője.
Oktatási tevékenységet a BME Villamosmérnöki Karán infokommunikációs szabályozás témában, valamint a Pécsi Tudományegyetem Állam és Jogtudományi Karának posztgraduális infokommunikációs szakjogász képzésén vezetékes és vezeték nélküli hírközlés témában folytat. 2006 óta a BME címzetes egyetemi docense.
A HTE-nek 1978 óta tagja, 1990 óta vesz részt különböző pozíciókban a vezetésében, 2011–2017 között a HTE főtitkára volt. A Híradástechnika folyóiratnak 1990-től 2011-ig volt szerkesztőbizottsági tagja. A HTE MediaNet és HTE Infokom konferenciák szervezésében programbizottsági tagként, szekcióvezetőként és előadóként is rendszeresen részt vesz. A hazai szakmai konferenciák állandó előadója, számos publikáció, tanulmány és előadás szerzője és több szakkönyv társszerzője, szerkesztője. Az új technológiák mellett hobbiként a távközléstörténettel is foglalkozik, írásai a HTE honlapján jelennek meg.

Felhők biztonsági kérdéseinek aktualitásai

KOVÁCS ZOLTÁN
Vodafone Magyarország Zrt.
zkovacs.24@gmail.com

Kulcsszavak: felhő, mesterséges intelligencia, ellátási lánc, kiberbiztonság

Jelen cikk rávilágít a felhőalapú rendszerek megkerülhetetlenségére, példákat hoz a speciálisan a felhőalapú rendszereknél jelentkező biztonsági problémákra, majd bemutat egy lehetséges módszertant annak eldöntésére, hogy egy felhőalapú rendszer teljesíti-e a minimálisan elvárt biztonsági szintet. A biztonság szemszögéből felvázolja a privát és az állami szférában használt felhőalapú rendszerek lényegesebb különbségeit, valamint példákon keresztül mutatja be az ellátási lánc védelmének jelentőségét. Példákat hoz a mesterséges intelligencia, és annak felhőben működő megoldásainak védelmi célú használatára, amely mellett bemutatja annak támadó oldalon történő felhasználást is.

1. Bevezetés

Napjainkban a felhőalapú rendszerek használata megkerülhetetlen. Ma már állami szervezetek is egyre gyakrabban használnak ilyen rendszereket, kihasználva annak költséghatékonyságát és egyéb előnyös tulajdonságait. Ez Magyarországra is igaz. Ugyanakkor a felhőalapú rendszerek használatakor bizonyos, több esetben a teljesen saját infokommunikációs rendszertől eltérő biztonsági kihívással is számolniuk kell a felhasználóknak. Ezeket pedig a kockázatelemzésük során figyelembe kell venniük.

A felhőalapú rendszerek biztonsága az elmúlt években sokat fejlődött. Sok esetben akár biztonságosabb is lehet a felhasználó számára, mint egy teljesen saját rendszer használata. Annak elemzésére, hogy egy felhőalapú rendszer biztonsági szempontból mennyire felel meg egy adott szervezet elvárásainak, ma már megvannak a módszerek.

Mindezek mellett egy olyan jelenség is megfigyelhető, hogy épp a felhőalapú rendszerek használatával egyre több kiszervezés, majd onnan bizonyos feladatok további kiszervezése, azaz újabb felek bevonása valósul meg. Így a korábbihoz képest is újabb ellátási láncok jönnek létre, az egyes szervezetek pedig egyre inkább csak az alaptevékenységükre koncentrálnak. Ez azonban új veszélyeket is rejt, amelyekkel a jövőben számolniuk kell.

Jelen cikk az első szakaszában előrejelzéseken keresztül világít rá a felhőalapú rendszerek megkerülhetetlenségére, példákat hoz a speciálisan a felhőalapú rendszereknél jelentkező biztonsági problémákra, majd bemutat egy lehetséges módszertant annak eldöntésére, hogy egy felhőalapú rendszer teljesíti-e a szervezet által minimálisan elvárt biztonsági szintet. A második szakasz biztonsági aspektusból felvázolja a privát és az állami szférában használt felhőalapú rendszerek lényegesebb különbségeit, valamint példákon keresztül mutatja be az ellátási lánc védelmének jelentőségét. A harmadik szakasz a mesterséges intelligencia védelmi célú használatára hoz példákat.

2. A felhőalapú rendszerek megkerülhetetlensége és biztonsága

A felhőalapú rendszerek használata egyre jobban terjed. Érezzük ezt mindennapjainkban is, hiszen magánemberként is egyre több ilyen rendszert használunk. Igénybe vehetünk banki szolgáltatásokat és fizethetünk webáruházakban, játszhatunk, szerkeszthetjük dokumentumainkat, képeinket, tárolhatjuk, megoszthatjuk adatainkat, készíthetünk útvonaltervet, és még nagyon hosszan lehetne folytatni a felsorolást. Igaz ez a vállalati felhasználásra is. Az elmúlt években egyre nőtt és az előrejelzések szerint várhatóan a jövőben is tovább növekszik a felhőalapú rendszerek felhasználása. A Canalys cég elemzése szerint a felhő-infrastruktúrára költött összegek világviszonylatban negyedévről negyedévre folyamatosan nőttek, és bár a növekedés üteme kicsit lassult, az még így is 35% körül mozog [1]. Ezt mutatja be az *1. ábra*.

A Technavio szerint a növekedés valóban lassul, de 2025-ig szóló előrejelzésük szerint a globális felhőpiac átlagos éves növekedési üteme 2021–2025 között még így is 17% felett várható. A növekedéshez egyedül Észak-Amerika mintegy 40%-kal járul majd hozzá [2]. A Grand View Research 2028-ig szóló előrejelzésében a növekedés éves átlagos ütemét globálisan 19,1%-ra becsüli, az Egyesült Államok piacán pedig 18,1%-ra úgy, hogy a piac megoszlása a szolgáltatási modellek (infrastruktúra mint szolgáltatás, platform mint szolgáltatás és szoftver mint szolgáltatás) között arányaiban nem változik jelentősen [3].

A fentiekből látszik, hogy a felhőalapú rendszerek felhasználása a közeljövőben is erőteljesen növekedni fog.

A felhasználásnál azonban a biztonságra, különös tekintettel a felhőalapú rendszereknél jelentkező biztonsági kihívásokra is figyelemmel kell lenni. Ezek azok a kérdések, amelyek egy teljesen saját infokommunikációs rendszer esetében lényegesen más hangsúlyokkal jelennek meg, vagy adott esetben egyáltalán nem jelennek meg. Kiragadott példák ilyenek lehetnek:

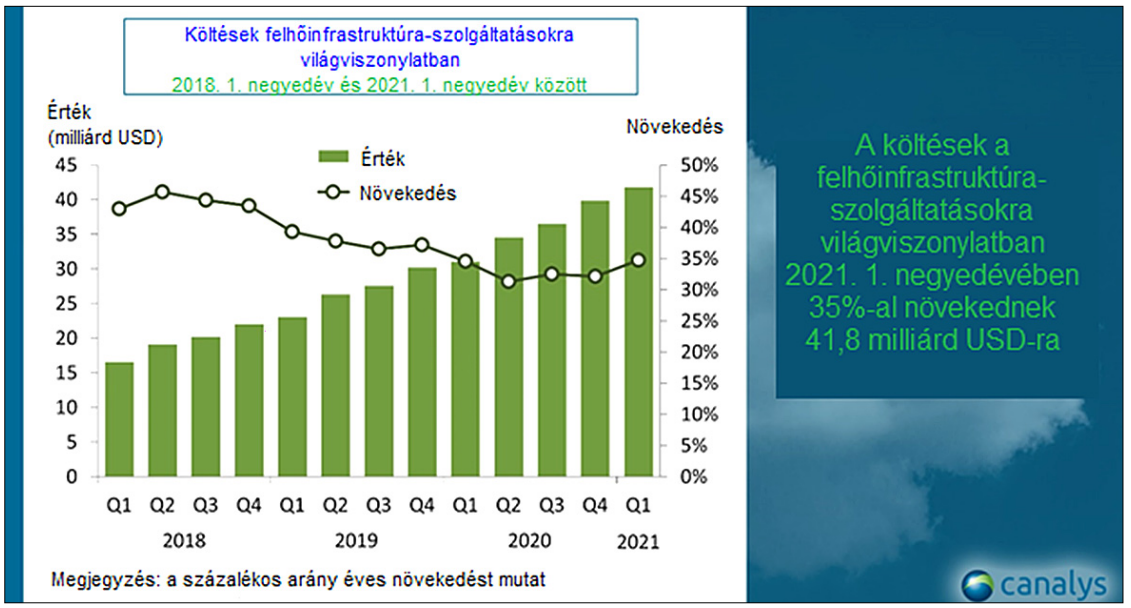
- *adatokkal kapcsolatos biztonsági meggondolások* (adatok, naplóadatok tulajdonjoga, adatok migrációja, adatok biztonsága és adatvédelemi kérdések, biztonsági tesztelések célja és hatálya, hibajavítások, adatmegőrzés és -törlés, interoperabilitás stb.);
- *adatmegőrzési kérdések* (adott esetben a jogszabályi kötelezettségekből adódó adatmegőrzés hogyan biztosítható, az adataink valóban törlésre kerüljenek, azok ne legyen visszaállíthatók, még a back-up-ból sem stb.);
- *a kínált és szükséges biztonsági intézkedések* (hozzáférési pontok, biztonsági értékelések és jogosultságok, folyamatos ellenőrzés, biztonsági ellenőrzések implementálása stb.);
- a felhőben használt *virtuális operációs rendszerek biztonsága* (többfelhasználós környezet problémái, izoláció stb.);
- *titkosítás* szükségessége (a felhő használatához, a felhő-erőforrások felügyeletét szabályozó interfész, operációs rendszerek, alkalmazások, adatok eléréséhez);
- *naplózási* kérdések (mi kerül naplózásra, ki és hogyan férhet hozzá a naplóadatokhoz stb.);
- *szolgáltató kémkedésének* kérdésköre;
- *harmadik fél bevonása* (mikor és mihez férhetnek hozzá a szolgáltató olyan alvállalkozói, akikkel a felhasználónak nincs szerződése stb.).

A fentiekből látszik, hogy a kockázatelemzést és a kockázatok csökkentésére szóló intézkedéseket más hangsúlyokkal kell elvégezni, mint egy tisztán saját rendszer esetében, és a szolgáltatóval való együttműködés, adott kérdések szerződésben való rögzítése pedig elengedhetetlen. Annak érdekében, hogy megvizsgálhassuk, hogy egy felhőalapú rendszer és annak szolgáltatója megfelel-e a szervezet által támasztott biztonsági követelményeknek, több módszert is alkalmazhatunk.

Az egyik ilyen lehetőség, hogy megvizsgáljuk az alábbi kérdéseket:

Üzembiztonság
- – közösségi felhő közös üzembiztonsági alapkövetelményei
- – hordozhatóság és interoperabilitás
- – üzletmenet-folytonosság
- – rendelkezésre állás, ellenálló-képesség, megbízhatóság
- – szolgáltatási szintek, azok garanciái
- – skálázhatóság és erőforrás-menedzsment
- – redundancia
- – katasztrófa-elhárítási terv
- – biztonsági mentés és visszaállítás
- – életciklus-kezelés és változásmenedzsment
- – adatmigráció
- – adatformátum
- – adatközpont működése, rendszerkonfiguráció
- – alkalmazásbiztonság
- – javítócsomagok kezelése
- – ellátási lánc üzembiztonsága
- – üzembiztonsági kockázatcsökkentés

Adatbiztonság
- – közösségi felhő közös üzembiztonsági alapkövetelményei
- – irányítás
- – kockázatkezelés
- – azonosítás, hitelesítés jogosultságkezelés és hozzáférés-szabályozás
- – határvédelmi eszközök működtetése és ellenőrzése
- – folyamatos ellenőrzés
- – incidenskezelés
- – sérülékenység vizsgálata, kezelése
- – titkosítás és kulcskezelés



1. ábra
A felhő-infrastruktúrára költött összegek változása világviszonylatban.
(Forrás: https://canalys.com/newsroom/global-cloud-market-Q121)

– virtualizációból adódó biztonság kezelése
– kliensoldali védelem
– vezeték nélküli hálózatok biztonsága
– biztonsági architektúra
– adatok elvesztése, ellopása
– mobil eszközök kezelésének terve

Egyéb (jogi, fizikai stb.) biztonság

– jogi megfelelőség
– fizikai biztonság
– személyi biztonság
– gazdasági biztonság
– dokumentumbiztonság

Azért ezen kategóriákon keresztül és nem a szokásos bizalmasság-sértetlenség-rendelkezésreállás mentén érdemes foglalkozni a kérdéskörrel, mert egyrészt az egyes szervezeteknél különböző emberek, jobb esetben csoportok vagy szervezeti egységek foglalkoznak a fent említett témakörökkel, másrészt a szakirodalomban is találkozhatunk hasonló megközelítésekkel, így például felosztásában közel azonos jellegűt lehet találni a BSI [4] és szinte teljesen megegyezőt az ENISA [5] anyagaiban.

Az elemzéshez használható például az alább röviden ismertetésre kerülő elemző sablon, amely a felhőalapú rendszer minimálisan elvárt biztonsági szintjének megállapítására szolgál. Ebből mutat be egy részletet a *2. ábra*.

A sablon az említett három főcsoportra bomlik, az egyes főcsoportokon belül további alcsoportok találhatóak. Szürke színnel a kiemelt fő vagy összegző, míg fehérrel az egyes részletkérdések találhatók meg. A kérdésekhez igen/nem válaszok adhatók, a fehér kérdések közül a felhasználó döntése alapján egy vagy akár több is elhagyható, a szürkék kitöltése kötelező. A fehérrel jelölt alcsoporti kérdések esetében a nemleges válasz még nem jelenti automatikusan azt, hogy a vizsgált felhőalapú rendszer nem felel meg a megkívánt biztonsági követelményeknek. A szürkével jelölt összegző, vagy fő kérdések esetében a nemleges válasz viszont igen [6]. Egy ilyen elemző sablon kitöltése után a szervezet el tudja dönteni, hogy az adott felhőalapú szolgáltatás és annak szolgáltatója megfelel-e az általa elvárt minimális biztonsági követelményeknek. Amennyiben igen, akkor az ezen felüli részletes biztonsági követelményeket a későbbiekben tisztázni, majd a szerződésben rögzíteni tudják.

Az elemzéshez más, például a FedRAMP [7], vagy az ott leírtakhoz hasonló sablon is használható.

2. ábra Részlet a biztonsági elemző sablonból.

Kategóriák vizsgálandó kérdései	válaszok	
	igen	nem
Üzembiztonság:		
1. közösségi felhő közös üzembiztonsági alapkövetelményei		
A felhasználó számára megfelelő és elfogadható a közösségi felhő felhasználói által kialakított közös üzembiztonsági alapkövetelmény rendszer?	☐	☐
2. hordozhatóság és interoperabilitás		
A szolgáltató által kínált eszközök, rendszerek és szolgáltatások biztosítják az adatok hordozhatóságát és a különböző (szolgáltató, felhasználó, harmadik fél) rendszereinek együttműködését?	☐	☐
A szolgáltató által kínált eszközök, rendszerek és szolgáltatások biztosítják az együttműködési képességet más, saját országbeli érintett hatóság vagy közigazgatási szervvel?	☐	☐
A szolgáltató által kínált eszközök, rendszerek és szolgáltatások biztosítják az együttműködési képességet más országok vagy EU érintett rendvédelmi szerveinek rendszereivel?	☐	☐
◦ A szolgáltató nyilvános, publikált API-kat használ?	☐	☐
◦ A szolgáltató által kínált interfészek interoperabilisak		
▪ a felhasználó rendszereinek interfészeivel?	☐	☐
▪ a felhasználó munkafolyamataihoz kapcsolódó harmadik felek rendszereinek interfészeivel?	☐	☐

3. Az ellátási lánc védelmének jelentősége

Mielőtt rátérnénk az ellátási lánc védelmének kérdéseire, érdemes áttekinteni, mit jelent egy kormányzati felhő használata biztonság szempontjából. Egyrészt azért, mert mások lehetnek a biztonsági elvárások, az állami szervezetek által használt felhőalapú rendszereknél sokszor vagy szigorúbbak, vagy – a jogszabályi előírások miatt – kötöttebbek. Másrészt pedig azért, mert hazánkban is használnak felhőalapú megoldásokat a kormányzati szervek.

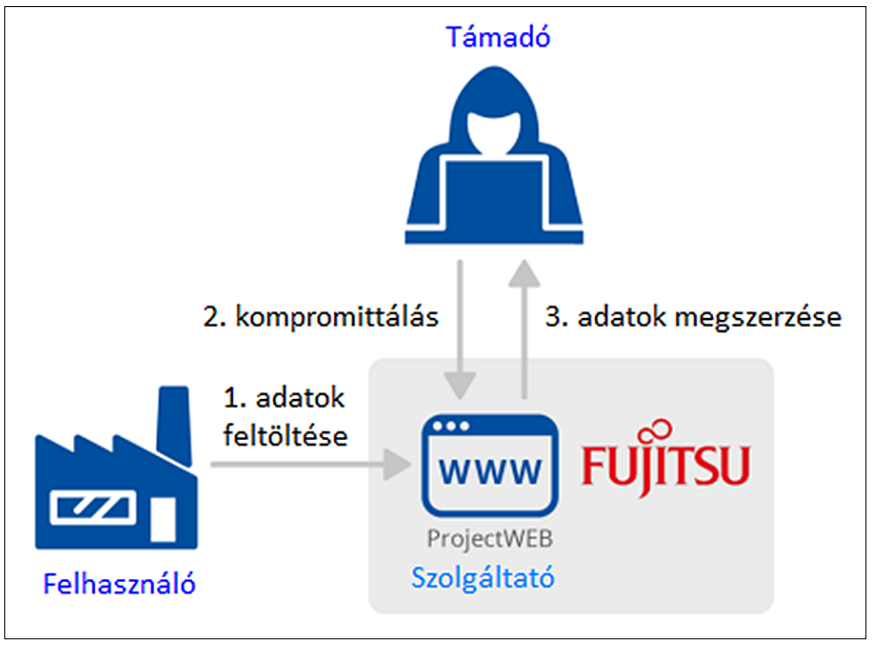
Magyarországon a NISZ Zrt. nyújtja a kormányzatifelhő-megoldásokat, így hazánkban – előremutató módon – állami szolgáltatás keretében biztosítanak meghatározott szervezetek számára felhőalapú rendszereket. Ezzel sikerül kihasználni a felhő előnyeit (pl. költséghatékonyság, megfelelő színvonalú hardver- és szoftverelemek, megfelelő mennyiségű és tudású szakembergárda stb.) és azokat ötvözni egy magasabb szintű biztonsággal. Azzal ugyanis, hogy a szolgáltató hazai, ráadásul állami tulajdonú is, a biztonsági problémák jelentős részét megoldották. A korábban a felhőalapú rendszerek kapcsán jelzett biztonsági felvetéseket ugyanis egyrészt jogszabályi alapon lehetett kezelni, másrészt az államnak minden lehetősége megvan ezen előírások betartatására és azok ellenőrzésére. Ez pedig olyan kérdésekben is megnyugtató választ tud adni, mint a lentebb ismertetésre kerülő ellátási lánc elemei esetében a megfelelő biztonság kialakítása.

Ugyanakkor meg kell jegyezni, hogy így is maradtak fenn kihívások. Az egyik, hogy a kormányzati felhőt meghatározott szervezeti kör használhatja. Ez természetes, mert ez a deklarált célja. Ugyanakkor ez azt is jelenti, hogy nem minden állami, önkormányzati szerv tudja használni. Nekik viszont vagy saját infokommunikációs rendszert, vagy – amennyiben ez lehetséges – publikus felhőt kell használniuk. A másik, hogy egy kormányzati célú felhőalapú rendszer továbbfejlesztése mindig korlátozottabb, mint egy privát rendszer esetében. Harmadrészt a kormányzati felhő felhasználói „shadow IT"-jelleggel használhatnak úgy privátfelhő-megoldásokat, hogy az nem engedélyezett, vezetőiknek arról nincs is tudomása. Például levelezést folytathatnak Gmail-en keresztül, vagy ezen keresztül küldenek át maguknak olyan anyagokat, amelyekkel az otthoni gépükön is dolgoznak egy szoros határidő miatt. Ezek mind olyan biztonsági kihívások, amelyekkel a szakembereknek foglalkozniuk kell.

A felhőalapú rendszerek egyre nagyobb arányú használata az ellátási láncokra is jelentős hatást gyakorolt. A korábban az egyes szervezetek által tulajdonolt és üzemeltetett rendszerek részben vagy egészben máshoz kerültek, bizonyos feladatokkal együtt. Ezzel új elemek vagy ágak jelentek meg az ellátási láncokban. Ugyanakkor biztonsági szempontból ennek veszélyei is vannak. Az ENISA 2021-ben kiadott tanulmánya [8] is erre hívja fel a figyelmet. Tanulmányukban 24 olyan publikusan is bejelentett és igazolt, az ellátási láncot ért támadást elemeznek, amelyek 2020. január és 2021. július között következtek be.

Ezek közül egy kiragadott példa a Fujitsu Projectweb elleni támadás. A Projectweb egy felhőalapú szoftver, amelyet a felhasználó szervezetek online együttműködésre, szoftvermenedzsmentre és fájlok megosztására használnak, és a japán kormányzati szervek is előszeretettel használják. A támadás következtében a támadók hozzáfértek kormányzati adatokhoz és többek között a japán légiforgalmi irányító adatait is ellopták. A támadás lefolyását mutatja be a *3. ábra*.

3. ábra A Fujitsu Projectweb elleni támadásának lefolyása. (Forrás: https://www.enisa.europa.eu/publications/threat-landscape-for-supply-chain-attacks)



Az ENISA tanulmánya jól szemlélteti az ellátási lánc védelmének jelentőségét. Elemzésük szerint az incidensek kb. 58%-ánál a támadások az ügyféladatok (beleértve a személyazonosításra alkalmas adatokat) és a szellemi tulajdont tartalmazó eszközök ellen irányultak, és a támadások 66%-ánál a beszállítók nem tudták, vagy nem jelentették az incidenst. Ez utóbbi jól szemlélteti azt is, hogy az ellátási lánc szereplői között más a kiberbiztonsági, ezen belül az incidensek jelentésének érettségi szintje. Az ENISA tanulmánya szerint az ellátási láncot ért támadások 2021-ben várhatóan 4-szeresére nőnek 2020-hoz képest.

Biztonsági szempontból ezt a problémakört mindenképp tovább kell vizsgálnia azon cégeknek, akik ilyen kiszervezéseket végeztek, és/vagy felhőalapú rendszereket használnak. Ennek kapcsán olyan kérdéseket is vizsgálniuk kell, mint az, hogy a szolgáltató cég is tovább szerződhet, azaz bizonyos feladatokat, funkciókat más beszállítótól vehet igénybe, vagy adott esetben egy felhőalapú szolgáltatást használhat a szerződéses feladatai teljesítéségez. Ebben az esetben is azonban a teljes láncon keresztül biztosítani kell az adatokat birtokló cég kiberbiztonsági követelményeit. Meg kell határozni, hogy ki és miért felelős pontosan, hogyan lehet védekezni a teljes ellátási láncban, kinek milyen kiberbiztonsági követelményeket kell teljesítenie, mikor és ki felé milyen jelentési kötelezettség van incidens esetén, hogyan lehet auditálni az alvállalkozók alvállalkozóit stb.

De más vagy akár fordított kihívások is jelentkeznek az ellátási láncok kapcsán. Ilyen például az 5G-magánhálózatok esete. Ebben az esetben valamely szolgáltató 5G-magánhálózatot létesít a felhasználó telephelyén, a felhasználó pedig ezen keresztül akár a teljes robotizált termelését irányíthatja néhány felügyeletet ellátó ember segítségével.

Az 5G kapcsán az EU hathatós lépéseket tesz a biztonság megteremtése érdekében. Így kiadta az 5G-biztonsággal kapcsolatos uniós eszköztárat (EU toolbox on 5G Cybersecurity) [9], amelynek alapján az egyes nemzeti hatóságok szabályozhatják az 5G biztonsági kérdéseit. Ezek alapján szolgáltatói oldalról a megfelelő védelem biztosított lesz. Ebben az esetben a felhasználói oldalról rendkívül fontos a megfelelő szintű védelem kiépítése, hiszen a támadók adott esetben akár komoly – a vállalat profiljától függően akár még a környezetre is kiható – rombolást tudnak okozni.

4. Mesterséges intelligencia védelmi célú használata

Ma már a mesterséges intelligencia és a gépi tanulás azok közé a buzzwordök közé tartoznak, amelyeket lépten-nyomon hallhatunk. Ugyanakkor számos tudományos cikk, a témával foglalkozó fórum és blogbejegyzés foglalkozik a témával. Ma már elmondhatjuk, hogy mind a mesterséges intelligencia, mind a gépi tanulás a mindennapok részévé vált a kiberbiztonságban, és sajnos nem csak a védő, hanem a támadó oldalon is.

A védő oldalon megjelenő, mesterséges intelligenciával és gépi tanulással felvértezett eszközökről a gyártók is előszeretettel kommunikálnak, ám ezek hasznosságát, felhasználhatóságát a kibervédelemmel foglalkozó szakemberek is fel- és elismerik. Egy 2018-ban megjelent interjú szerint egy közepes méretű vállalat naponta kb. 200.000 biztonsági eseménnyel találkozik [10]. Ezek feldolgozása pedig csupán emberi erőforrással nem lehetséges. Egy 2021-ben megjelent tanulmány [11] szerint a mesterséges intelligencia használatának a kiberbiztonságban az alábbi fő előnyei vannak:

A mesterséges intelligencia:

1. *idővel egyre többet tanul,* így jól fogja ismerni az üzleti folyamatokat, a hálózati forgalmat, az ott zajló folyamatok normál viselkedését;
2. *azonosítja az ismeretlen fenyegetéseket,* amelyeket az ember sok esetben nem ismer fel;
3. *sok adatot képes kezelni,* így az egyre növekvő mennyiségű logot, riasztást, eseményt stb. is;
4. *jobb sérülékenységmenedzsmentet biztosít,* mert gyorsabb értékelést és felmérést biztosít;
5. *jobb általános biztonságot nyújt,* mert rengeteg támadási formát képes felismerni és priorizálni azokat;
6. *csökkenti az ismétlődő folyamatokat,* így fizikai és lelki terhelést is levesz a kiberbiztonsági szakemberekről;
7. *felgyorsítja az észlelési és válaszidőt,* ezáltal jelentősen növeli a hálózat biztonságát;
8. *biztonságosabb hitelesítést biztosít,* mert számos eszközt tud használni ehhez (pl. biometrikus azonosítás, CAPTCHA[1], brute force[2] felismerés).

A mesterséges intelligencia védő oldali használatával kapcsolatban a FireEye kiberbiztonsági cég 2018-as „FireEye Cyber Defense Summit 2018” című konferenciáján volt egy kerekasztal-beszélgetés. Ezen David Gunning, a DARPA[3] mesterséges intelligenciát kutató programjának menedzsere azt jósolta, hogy a kibervédelemben az ún. Tier1-szintű operátorokat öt éven belül kiváltja a mesterséges intelligencia. A Tier1-szintű operátorok azok a szakemberek, akik a riasztások, kiberbiztonsági események közvetlen és gyors feldolgozását végzik, vagy gyorsan lezárva az adott vizsgálatot (vagy azért mert ismert esemény, vagy azért mert téves, azaz fals positive riasztás volt), vagy ha az további vizsgálatot igényel, akkor továbbítják azokat az ún. Tier2-szintű elemzőknek. Bár az Egyesült Államokban a Tier1-operátorok feladatköre meglehetősen egyszerű és jórészt betanított feladatokból áll, azért ez a jóslás így is jól szemlélteti a mesterséges intelligencia fejlődését és előretörését napjainkban.

A felhőalapú rendszerek, a mesterséges intelligencia felhasználható a csalás elleni küzdelemben is. A több országban is működő nemzetközi vállalatok önmagukban is hatékonyan tudják használni a felhőalapon működő, mesterséges intelligenciát használó csalás elleni platformokat, hiszen olyan előnyöket tudnak kiaknázni, mint a szabályrendszerek egységesítése, a legjobb gyakorlatok beépítése, vagy egy újonnan felismert csalási formák elleni védekezés azonnali elterjesztése az összes leányvállalatnál stb. Ugyanakkor felmerül a kérdés, hogy ha ezek a rendszerek az adott ágazatban dolgozó vállalatok számára is hasonló előnyökkel járnának, miért nem épültek még ki? Ebben az esetben hasonló problémák jelentkeznek, mint a kiberbiztonság kapcsán a felismert fenyegetések egymás közötti megosztásakor. Ebben az esetben is a bizalom, a lehetséges veszteségek és az egymás közötti verseny jelentkeznek mint befolyásoló tényezők, amelyek jelentősen lassítják vagy akár el is lehetetleníthetik az ilyen ágazati platformok létrejöttét, működését. A költségek megosztása és a várható nyereség okán ugyanakkor érdemes lenne ilyeneket működtetni.

A mesterséges intelligencia természetesen a támadó oldalon is megjelenik. A támadók mesterséges intelligenciát – valamint annak részhalmazát, a gépi tanulást – is használják hatékonyabb, automatizáltabb, agresszívabb és összehangoltabb támadások indítására. Ezek segítségével ugyanis jobban fel tudják térképezni és meg tudják érteni, hogy a célpont-szervezetek hogyan védik a rendszereiket. Erre egy példa az ügyfélszolgálatokon is használt, mesterséges intelligenciával támogatott nyelvi feldolgozó eszközök, amelyek a támadók kezében kifinomultabb adathalász-levelek előállítását teszi lehetővé [12]. A támadók a támadás különböző szakaszaiban használják, használhatják a mesterséges intelligenciát. Így például az önállóan működő kártékony kód elkészítéséhez és bejuttatásához, a hálózaton belüli terjedéshez, a védelmi technológiák intelligens megkerüléséhez, a kisméretű és alacsony sebességű adatlopáshoz, vagy a felhasználó „élethű” szimulálásához stb. [13]. A Dark Web piacterein számos mesterséges intelligenciát és gépi tanulást használó hackereszközt kínálnak, és a kibertámadásokra egy ökoszisztéma épült fel, amelyben a támadás, mint szolgáltatást[4] is igénybe lehet venni [14]. A gyakorlati tapasztalatok is azt mutatják, hogy a támadók használják már ezeket a technológiákat.

1 CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart, azaz teljesen automatizált nyilvános Turing-teszt a számítógép és az ember megkülönböztetésére – egy olyan teszt, ami képes megkülönböztetni az emberi felhasználót a robottól (sz.géptől).

2 Brute force, azaz nyers erő. Ez egy titkosításokkal, jelszóvédelemmel szemben alkalmazott támadási módszer, amely során a támadó kipróbál minden lehetséges variációt a jelszóra.

3 DARPA (Defense Advanced Research Projects Agency), az Egyesült Államok Védelmi Minisztériumának kutatásokért felelős részlege.

4 cyber-attack-as-a-service (CAaaS)

A fentiek alapján napjainkban a mesterséges intelligencia használata a támadó és a védő oldalon is a gépek közötti, sőt az intelligencia-intelligencia közötti küzdelemről szól. A 2019-ben megrendezett RSA-konferencián, amely a világ legnagyobb kiberbiztonsági eseménye, az egyik előadásban élőben bemutatták a gyakorlatban is, milyen az, amikor mesterséges intelligenciával támadunk és mesterséges intelligenciával védekezünk. Bár a bemutató olyan volt, mint a 90-es években, amikor a sakkszoftvereket/robotokat játszatták egymással, azért néhány dologra így is ráirányította a figyelmet. Az első, hogy ma már mindkét oldalon demonstrálható módon működik a technológia. A második, hogy mindkét oldalon hatékonyan fel is lehet használni ezt. A harmadik pedig az, hogy a bemutatottnál fejlettebb, egy vagy néhány, de adott feladat(ok)ra fókuszáló technológiával rendelkezik mindkét oldal.

5. Összefoglalás

Összefoglalásként elmondható, hogy a felhőalapú rendszerek biztonsága sokat fejlődött az elmúlt években. Vannak azonban olyan biztonsági kérdések, amelyek vagy csak a felhőalapú rendszereknél jelennek meg, vagy ott nagyon hangsúlyosak, és ezeket a felhasználónak egy felhőalapú rendszer használatának tervezése és igénybevétele során a kockázatelemzésnél és -kezelésnél figyelembe kell vennie. Ugyanakkor a felhőalapú rendszerek és a mesterséges intelligencia terjedésével, a kiszervezések bővülésével új kockázatok is megjelennek.

Az ellátási lánc támadása korábban nem látott méreteket ölt, így az adatokat tulajdonló szervezet elemi érdeke, hogy a biztonság az ellátási lánc teljes egészében a rá vonatkozó szabályok szerint biztosított legyen. Ez pedig a korábbiaknál részletesebb, mélyebb kockázatelemzést és -kezelést igényel a felhasználóktól. Ennek módszertana még nem teljes mértékben kidolgozott, ezzel a szakembereknek még foglalkozniuk kell.

Hivatkozások

[1] Canalys: Global cloud services market Q1 2021, 2020.04.29. (Letöltve: 2021.08.31.) *https://canalys.com/newsroom/global-cloud-market-Q121*

[2] Global Cloud Computing Market, Over $287 Billion growth expected during 2021–2025. Technavio. Cision. 2021.06.05. (Letöltve: 2021.08.31.) *https://www.prnewswire.com/news-releases/global-cloud-computing-market—over—287-billion-growth-expected-during-2021-2025—technavio-301306058.html*

[3] Cloud Computing Market Size, Share & Trends Analysis Report by Service (SaaS, IaaS), by Enterprise Size (Large Enterprises, SMEs), by End Use (BFSI, Manufacturing), by Deployment, and Segment Forecasts, 2021–2028. Market Analysis report. Grand View Research. 2021.07. (Letöltve: 2021.08.31.) *https://www.grandviewresearch.com/industry-analysis/cloud-computing-industry*

[4] Security Recommendations for Cloud Computing Providers (Minimum information security requirements), White Paper. 2011.06.22. (Letöltve: 2014.09.21.) *https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Minimum_information/SecurityRecommendationsCloudComputingProviders.html*

[5] Daniele Catteddu (szerk.): Security & Resilience in Governmental Clouds. 2011.01. (Letöltve: 2014.11.18.) *http://www.enisa.europa.eu/activities/risk-management/emerging-and-future-risk/deliverables/security-and-resilience-in-governmental-clouds*

[6] Kovács Zoltán: Az infokommunikációs rendszerek nemzetbiztonsági kihívásai. Doktori (PhD) Értekezés, Nemzeti Közszolgálati Egyetem, 2015.

[7] FedRAMP. (Letöltve: 2021.08.31.) *https://www.fedramp.gov/documents-templates/*

[8] ENISA Threat Landscape for Supply Chain Attacks. 2021. július. (Letöltve: 2021.08.31.) *https://www.enisa.europa.eu/publications/threat-landscape-for-supply-chain-attacks*

[9] Biztonságos 5G hálózatok: a Bizottság jóváhagy egy uniós eszköztárat, és meghatározza a következő lépéseket. 2020.01.29. (Letöltve: 2021.08.31.) *https://ec.europa.eu/commission/presscorner/detail/hu/ip_20_123*

[10] Dan Patterson: How AI, IoT, and big data will shape the future of cybersecurity. TechRepublic, 2018.08.13. (Letöltve: 2021.08.31.) *https://www.techrepublic.com/article/how-ai-iot-and-big-data-will-shape-the-future-of-cybersecurity/*

[11] Daniel Martin: 8 Benefits of Using AI for Cybersecurity. Cyber Management Alliance. 2021.05.04. (Letöltve: 2021.08.31.) *https://www.cm-alliance.com/cybersecurity-blog/8-benefits-of-using-ai-for-cybersecurity*

[12] Mercedes Cardona: When Bad Guys use AI and ML in Cyberattacks, What Do You Do? SecurityRoundtable.org, Powered by: Palo Alto Networks. (Letöltve: 2021.08.31.) *https://www.securityroundtable.org/when-bad-guys-use-ai-and-ml-in-cyberattacks-what-do-you-do/*

[13] DarkTrace: The Next Paradigm Shift AI-Driven Cyber-Attacks. Research White Paper, 2018. (Letöltve: 2021.08.31.) *https://www.oixio.ee/sites/default/files/the_next_paradigm_shift_-_ai_driven_cyber_attacks.pdf*

[14] Keman Huang, Michael Siegel, Keri Pearlson and Stuart Madnick: Casting the Dark Web in a New Light: A value-chain lens reveals a growing cyber attack ecosystem and new strategies for combating it. Massachusetts Institute of Technology, 2019. 06. (Letöltve: 2021.08.31.) *http://web.mit.edu/smadnick/www/wp/2019-19.pdf*

A szerzőről

KOVÁCS ZOLTÁN a Budapesti Műszaki Egyetemen szerzett diplomát a Villamosmérnöki Kar Híradástechnika szakán 1991-ben, ezt követően pedig 2004-ben mérnök-közgazdász diplomát a Budapesti Közgazdaságtudományi Egyetemen. PhD-értekezését 2015-ben védte meg a Nemzeti Közszolgálati Egyetemen. Dolgozott a Nemzetbiztonsági Szakszolgálatnál, a NISZ Zrt.-nél. Jelenleg a Vodafone Magyarország Zrt.-nél vezeti a biztonsági területet és tanít a Nemzeti Közszolgálati Egyetemen. Kutatási területe az infokommunikációs rendszerek biztonsági kihívásai.

Cloud-infrastruktúra és a rajta futó telekom-applikációk aktuális biztonsági kihívásai

CSORDÁS GÁBOR
Nokia
gabor.csordas@nokia.com

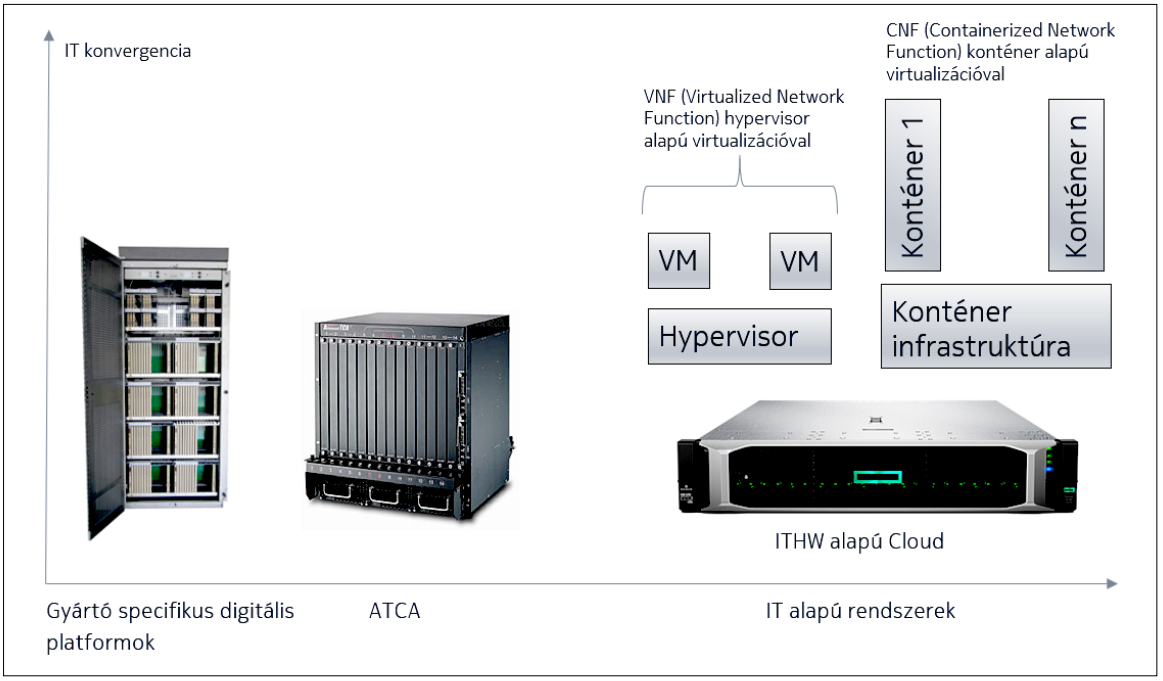
Kulcsszavak: cloud-infrastruktúra, nyílt forráskód, biztonság, sérülékenység, CVE

A távközlési iparágban használt infrastruktúra- és platform-megoldások egyre inkább konvergálnak az egyéb IT-megoldások által használt platformok irányába. Ez rengeteg előnnyel jár technikai és üzleti szempontból is, ugyanakkor olyan fenyegetések jelentek meg a távközlés területén is, amik eddig csak mérsékelten, vagy egyáltalán nem voltak jellemzőek. Mivel a telekommunikációs hálózatok kritikus rendszereknek tekintendők, különös figyelemmel kell kísérni ezen klasszikusan IT-eredetű, de immár a telekom-rendszereket is érintő fenyegetéseket. A teljesség igénye nélkül ide tartoznak a privát és publikus cloud-rendszerek általános sebezhetőségével kapcsolatos kérdések, a szabad vagy nyílt forráskódú szoftverek beépítése az infrastruktúrába és a telekom-applikációkba, illetve a privát cloudból a publikus cloud-rendszerek felé elmozdulás miatti koncepcionális változások.

1. Bevezetés

Elsődleges témánk a telekommunikációs hálózatok által használt infrastruktúrák bemutatása biztonsági szempontból, kitérve a legfőbb fenyegetési típusokra, amik manapság ezekre a platformokra leselkednek. Bár a cikk főleg technikai részletekre fókuszál, de a teljes kép kialakításához szükséges a technikai részletek mellett üzletmenetet érintő, folyamatirányítási és felelősséggel kapcsolatos kérdések tárgyalása is. Emellett érintőlegesen foglalkozni kell a telekommunikációs infrastruktúra részét nem képző, de annak biztonságát befolyásoló tényezőkkel is: magukkal a telekommunikációs applikációkkal, illetve a telekommunikációs szolgáltatásokat igénybe vevő eszközökkel.

Az írás első szakaszában a telekommunikációban használt infrastruktúra fejlődését tekintjük át röviden a biztonsági kihívások szempontjából, majd foglalkozunk az infrastruktúrát, illetve annak biztonságát érintő egyéb tényezőkkel. Ezt követően néhány komplex támadási típus bemutatására kerül sor: direkt az infrastruktúra elleni, vagy a megtámadott infrastruktúra felhasználásával további, tipikusan a telekommunikációs applikációk elleni támadásokhoz. Egy speciális veszélyforrás is részletesebben bemutatásra kerül, kifejtve miért jár veszélyekkel az OSS (Open Source Software) használata a telekommunikációs infrastruktúrában és applikációkban. Végül néhány, az infrastruktúra részét nem képző, de a hálózat biztonságát ettől még befolyásoló tényezőről is szükséges szót ejteni.



1. ábra A telekommunikációban használt infrastruktúra fejlődése.

2. A telekommunikációs infrastruktúra technikai fejlődésének történelmi áttekintése

A napjainkban a telekommunikációs infrastruktúrára leselkedő veszélyek megértéséhez szükséges áttekinteni, hogyan is jutottunk el a ma használt infrastruktúrákhoz, hogyan változott a fenyegetés jellege és mértéke ezen evolúció alatt. A későbbi szakaszok az ebben a szakaszban ismertetett fejlődési szakaszokra fognak hivatkozni a fenyegetettségek tárgyalásánál.

2.1. Kezdeti digitális megoldások

A telekommunikációs alkalmazások elterjedése digitális platformon a nyolcvanas évektől vált jelentőssé, amikor azok tipikusan valamilyen alkalmazás-specifikus HW- és SW-alapon futottak. Ahol lehetőség adódott rá, a gyártók igyekeztek valamilyen x86-alapú HW-t készíteni, de ezek attól még gyártó- és iparág-specifikusak maradtak. Az akkori teljesítménye az x86 architektúrának sok telekommunikációs alkalmazás esetén nem is tette lehetővé az arra való migrálást, ezért mindig is voltak olyan alkalmazások, amelyek inkább célorientált HW-en futottak, például valamilyen DSP-platformon. Ilyen DSP-n futó alkalmazások a mai napig léteznek, bár ezeknek a DSP-n tartása inkább üzleti döntésekre vezethető vissza. (A kifutó technológiákhoz tartozó alkalmazásokat tipikusan nem éri már meg migrálni a modernebb környezetbe.)

Az alkalmazások megírásához használt programnyelv is valamilyen az adott környezet és telekomspecifikus variáns volt (pl. gyártóspecifikus SDL-variánsok [1]), saját operációs rendszert használva.

Ezek a kezdetleges digitális platformok többszörös ráncfelvarráson estek keresztül, gyártói stratégiák függvényében 3G-alkalmazásoknál is még megtalálhatóak voltak a modernizált verzióik, de a ma futó telekommunikációs szolgáltatásokat nyújtó eszközök közül már teljesen eltűntek, a rajtuk futó releváns alkalmazásokat valamilyen modernebb platformra ültették át. Egy ilyen gyártóspecifikus digitális platformra jó példa a Nokia DX 200 [2].

2.2. x86-konvergencia, kezdetleges virtualizáció

Az x86-architektúra terjedésével megjelentek a legalább telekomon belüli platform-szabványosítási törekvések. Erre jó példa az ATCA-platform [3], sokszor itt már valamilyen Linux-alapú, de telekom-igényekre optimalizált operációs rendszert és kezdetleges hypervisor-alapú virtualizációs technológiákat használva. Ezek a megoldások már tekinthetőek a cloud-alapú infrastruktúra előfutárainak, de attól még jelentősen különböztek az alábbi pontokban:

• A használt ATCA HW ugyan a szabvány szerinti volt, de a legtöbb gyártó alkalmazásait csak saját ATCA HW-en támogatta. Tehát volt ugyan egy elvileg közös szabvány, amit a gyártók közösen próbáltak követni, de a szolgáltató szempontjából továbbra is összeforrott egység formájában történt a HW és rajta futó alkalmazás kiszállítása, a különböző gyártók elvileg például ATCA-specifikus komponenseit nem lehetett keverni, az egyik gyártó alkalmazását nem lehetett egy másik gyártó ATCA HW-környezetben futtatni a gyártói támogatás megtartása mellett.

• Amennyiben a használt operációs rendszer valamilyen Linux-alapra épült, rengeteg módosítást tartalmazott a telekom által megkövetelt speciális rendelkezésre állás, redundancia és valós idejű feldolgozási igények miatt. Az operációs rendszer, ide értve az esetleges hypervisort is, és a rajta futó alkalmazás elválaszthatatlan egységet képzett.

• A hypervisor-alapú virtualizáció itt még igen kezdetleges volt, mivel az többnyire statikus módon történt, az applikációt felépítő virtuális gépek leosztása többnyire előre definiált (az adott operátorra vonatkozóan), sokszor speciális igényű virtuális gépekkel, ami miatt a platform nem is mindig volt teljesen homogén.

2.3. IT-konvergencia

A következő logikus lépés a „telekom siló” felszámolása lett, cél az IT-területen már széles körűen és költséghatékonyan használt HW- és SW-komponensek integrálása:

– blade/rackmount alapú szabványos HW, ami már nem feltétlenül a telekom-beszállítótól kell származzon (IBM, Dell, HPE, ...);
– már széles körűen elterjedt operációs rendszerek használata, vagy legalább azok alapul vétele, amire aztán a beszállító elkészíttette a szükséges módosításokat (Ubuntu, Red Hat, ...);
– teljes hypervisor-alapú virtualizáció, a hozzá tartozó management- és kiegészítő funkciókkal (VMware NFV, OpenStack...).

A telekom-applikációknak azonban ezen a már szinte teljesen IT-alapú környezeten is szükségük van egy speciális VNF- (Virtualized Network Function) managementszintre, a legtöbb gyártó az ETSI MANO-t [4] követte ezen a területen. Mindezek mellett a speciális rendelkezésreállási és valósidejű feldolgozási igények miatt az operációs rendszer és a cloud SW szintjén továbbra is szükség volt apróbb módosításokra, amit vagy a beszállító saját maga végzett el, vagy partneri kapcsolatban a cloud SW beszállítóival közösen vittek végbe.

2.4. Cloud native infrastruktúra

A konténeralapú virtualizáció megjelenésével a telekommunikációs alkalmazások még közelebb kerültek az infrastruktúra felhasználása szempontjából az IT-iparághoz:

– szabványos IT HW,
– széles körben használt operációs rendszerek és cloud-infrastruktúra.

Az 5G-alkalmazások jelentős része már ebbe a környezetbe született bele, ezért sokkal könnyebb volt azoknak az implementálása ilyen módon úgy, hogy a cloud-infrastruktúra felé ne kelljen speciális telekom-jellegű igényeket támasztani, amiket azok a megfelelő tervezéssel és méretezéssel ne tudnának kielégíteni.

A beszállítók termékei között megtalálhatóak ugyan saját operációs rendszer és cloud SW-termékek is (ezek többnyire valamelyik, a piacon elérhető megoldás továbbgondolásai), de a fő hangsúly a telekom-applikáción van, a gyártók portfoliójában egyre kisebb szeletet képvisel a HW- és a SW-infrastruktúra. Amennyiben a beszállító portfoliójában vannak ilyen termékek, azokat már tipikusan nem csak a saját applikációihoz kínálja, hanem akár önmagában is, és mint HW- vagy teljes cloud-infrastruktúra megoldást árulja, ebben az esetben a klasszikus IT-beszállítók konkurenciája szerepkörben.

3. Üzleti folyamatok és percepciók változása, a fenyegetettség szintjének és jellegének változása az infrastruktúra fejlődése során

A technológia evolúciója mellett vagy részben annak a hatására is jelentős változások történtek a telekom-applikációk üzemeltetése terén is.

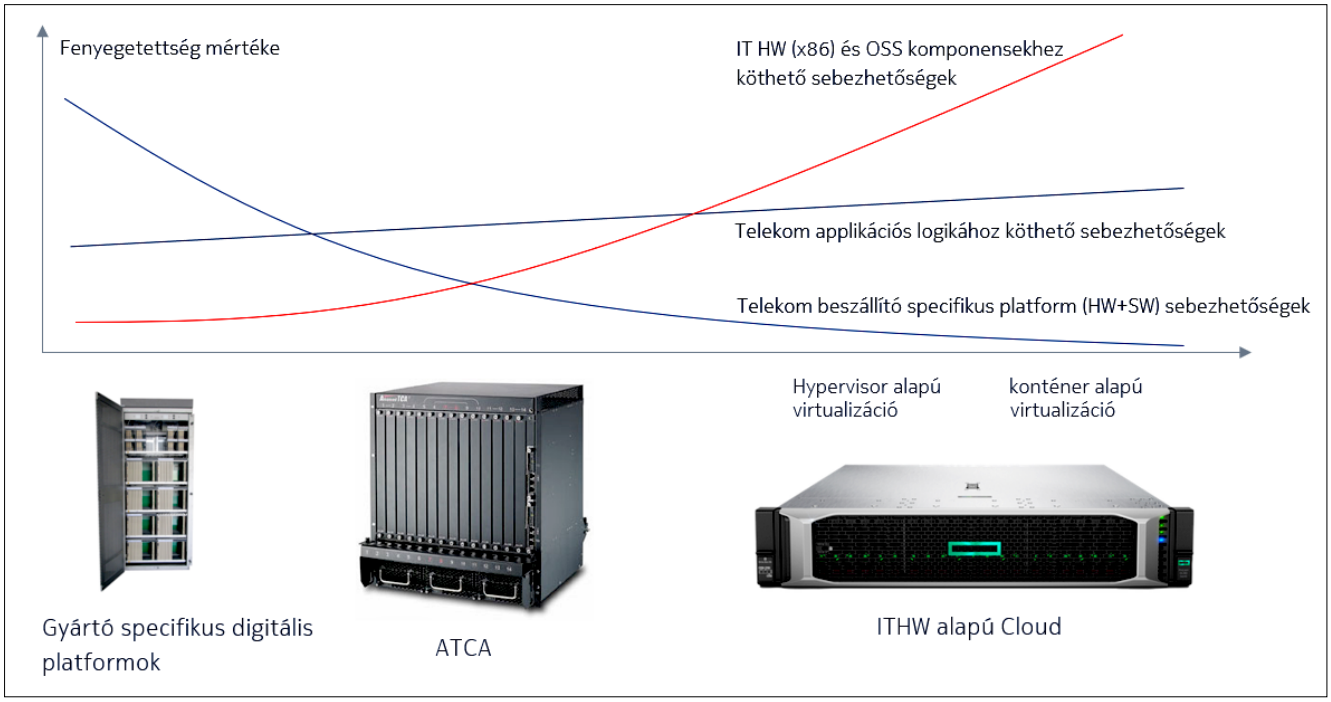
A korai időszakban az adott 2G-, 3G- és részben még a 4G-megoldásokat felépítő funkciók, vagy legalább a különböző alrendszerek egy beszállítótól származtak (pl. rádiós hálózat, core network, hálózatfelügyelet). Ennek köszönhetően üzemeltetési szempontból is azokra egy egységként tekintett mindenki, miszerint elég annak a határait megfelelően levédeni, azon belül egy DMZ-t (demilitarized zone) létrehozni, ahol már enyhébb követelmények vonatkoznak a biztonságra. Ne feledjük, hogy sok biztonsági követelménynek nagyon komoly kihatása lehet a teljesítményre is, tehát egy ilyen DMZ használata nem tekintendő könnyelműségnek, egyszerűen csak ez volt a beszállító és az üzemeltető közös érdeke az adott rendszer teljesítményének maximalizálása érdekében.

Az egyre komolyabb IT-irányú konvergenciának, egyre nyíltabb telekom szabványoknak és a növekvő komplexitásnak köszönhetően napjainkra ez a fajta gyártóhomogén környezet eltűnőben van. A DMZ határvonalai elmosódnak, egyre több gyártó egyre több terméke szükséges egy telekommunikációs hálózat megvalósításához. (Érdemes összevetni hány hálózati elem volt szükséges egy 2G-hálózat felépítéséhez és hány elemből épül fel egy 5G-hálózat.)

A határok védelme már nem elégséges, minden egyes funkciót vagy elemet külön, önmagában is fel kell ruházni az összes védelmi funkcióval. Fontos megjegyezni, hogy ebben a kontextusban a DMZ nem csak az elemek közötti kommunikációra vonatkozik, hanem a SW készítésének módjára is. Tehát nem csak arról van szó, hogy például IP szinten minden egyes elemet már önmagában is meg kell védeni, hanem azt is kontrollálni kell, hogy az adott elem fejlesztése, kiszállítása, telepítése és üzemeltetése során azzal mi történik: milyen beszállítói láncon keresztül milyen idegen komponensek kerülhetnek bele (lásd később az OSS-el foglalkozó szakaszt).

Fontos szerepe van az új technológiákhoz kapcsolódó percepcióknak is. Egy új technológia, mint például a konténeralapú virtualizáció megjelenésekor nem mindig annak a biztonságos üzemeltethetősége a legelső fő szempont. E cikk nem tér ki a hypervisor- és konténeralapú virtualizáció összehasonlítására, de általánosan elmondható, hogy a döntéshozók szemszögéből a konténeralapú virtualizáció jelenleg kevésbé tekintett biztonságosnak, mint a hypervisor-alapú, ezt több országspecifikus szabványban is nyomatékosították. Amikor egy infrastruktúra biztonságosságáról beszélünk, akkor az ilyen percepciókkal is foglalkozni kell, hiszen okkal vagy ok nélkül, de ezek is alakítják a biztonsággal kapcsolatos elvárásokat és irányelveket.

2. ábra A fenyegetettség változása a különböző infrastruktúrák esetében.



A fenti áttekintés után a technológia evolúció, üzletmenet és percepciók terén el is érkeztünk ahhoz a ponthoz, amikor ezek függvényében vizsgálható a rendszerre ható fenyegetettségi szint *(lásd a 2. ábrát)*.

Az első fontos észrevétel, miszerint a silóalapú kezdetleges, IT-tól távol álló infrastruktúra kevesebb, de legalábbis más jellegű veszélynek volt kitéve. Az IT felé történő migrációval a telekommunikációs iparág megörökölte az összes IT-sebezhetőséget is minden logikai szinten. Egy DSP-alapú platformra aránylag kevés támadási módszer ismert annak egyedisége és elszigeteltsége miatt, míg egy általános cloud-alapú infrastruktúra rengeteg módon támadható, jól dokumentált módszerekkel. Az IT-alapú rendszerekre már magasabb az infrastruktúrára leselkedő fenyegetési szint, ugyanakkor ezek jellege is megváltozott:

– a beszállító-specifikus HW-rel együtt az ahhoz kapcsolódó specifikus veszélyforrások is eltűnnek,
– az IT-alapú HW magával hozta az általános HW-sérülékenységeket (lásd a Spectre példáját az 5. szakaszban),
– a nagyfokú OSS- (Open Source Software) használat miatt azoknak a karbantartása a telekommunikációs alkalmazásokban egy soha véget nem érő küzdelemmé vált (lásd 6. szakasz).

Meg kell említeni a telekommunikációs applikációs logikát érintő fenyegetéseket is, amik a használt platformtól függetlenül jelen vannak. Mivel a telekommunikációs applikációk komplexitása növekszik, ezért ez a fenyegetettség is enyhén növekvő szinttel van jelölve a diagramon. (Például signaling protokollok sebezhetősége, részletesebben kifejtve az 5. szakaszban.)

(A diagramon a fenyegetési típusok trendje a fontos, nem feltétlenül azok egymáshoz viszonyított nagysága.)

4. Infrastruktúrát és az applikációs szintet érintő sérülékenységek

Mielőtt tovább elemeznénk az infrastruktúra támadhatóságát és annak főbb okait, szükséges jobban elkülöníteni az infrastruktúrára és az applikációra leselkedő veszélyeket, összehasonlítani és foglalkozni azzal, hogyan épül fel egy hatékony támadás, azt néhány példával illusztrálva. Mivel az infrastruktúra megörökölte az összes IT-sebezhetőséget, nem meglepő, hogy az ismert támadási metódusok jelentős része az infrastruktúrán keresztül indul (később esetlegesen behatolva az applikációs szintre is). Ennek főbb okai:

- Hatalmas installált bázis nagyon hasonló vagy teljesen egyforma HW- és SW-komponensekből felépítve.
- Jól dokumentált gyengeségek és támadási metódusok (pl. CVE [5] adatbázis).
- Toolkitek elérhetőek az ismert infrastruktúra sérülékenységek kihasználására, estenként ezeknek a kiaknázása még különösebben mély ismereteket sem igényel. Többször is előfordult már, hogy CVE formájában már elérhető volt egy sérülékenység leírása, annak a kihasználására már az automatikus eszközök is megjelentek azonnal, miközben az érintett sérülékeny komponens beszállítója még dolgozott a javításokon. Ez tulajdonképpen nyílt felhívás volt ezeknek a kihasználására ezen átmeneti időszak alatt, amíg különböző szinten átmeneti intézkedéseket kellett hozni a védelem érdekében.
- A később felhozott példák kapcsán látszani fog, hogy egy sikeres átfogó támadás esetében az infrastruktúra megsebzése (legyen az a CIA-hármas [6] bármelyik aspektusa) csak az első lépés. A tényleges applikációs logikát érintő támadások sokszor valamilyen infrastruktúrát érintő támadásra épülnek.

Az applikációs szint megtámadása már tipikusan jóval komplexebb feladat, sokkal mélyebb és specifikusabb tudást és eszközöket igényel:

- A telekommunikációs applikációk logikáját érintő sérülékenységek nincsenek annyira széleskörűen és nyíltan dokumentálva, mint a széles körben használt IT-rendszerek infrastruktúra-szintű sérülékenységei.
- A telekom-applikációk sikeres megtámadásához (ismét legyen az a CIA-hármas bármelyik aspektusa) sokkal részletesebb ismerete szükséges a távközlési szabványoknak, beszállító-specifikus megoldásoknak. (Signalling protokollok, ETSI-szabványok [7], architektúrák stb.)

5. Példák a támadások összetettségére

Ez a fejezet példákon keresztül mutatja be, hogyan is néz ki néhány tipikus támadási módszer. Nem foglalkozunk az olyan „egyszerű” módszerekkel, mint root-jogosultságok szerzése a jelszó vagy privát kulcs megszerzésével. A cél néhány ezeknél jóval összetettebb példa rövid bemutatása és annak illusztrációja, hogy egy sikeres támadáshoz többnyire nem elég egy adott sérülékenységet kihasználni, hanem arra valamilyen workflow-t kell építeni a kívánt cél elérése érdekében (attól függően, a CIA-hármas melyik eleme a cél).

5.1. HW-sérülékenység kihasználása

Egy HW-sérülékenység kihasználása önmagában még csak az első lépés szokott lenni. Jó példa erre a Spectre&Meltdown [8], illetve az L1TF [9] sérülékenység. Ezekkel a processzor regisztereiből, illetve cache-ből szedegethető össze információtöredék, amit felhasználva lehet folyatni a támadást. A Spectre esetében azonban ez egy rendkívül komplex folyamat.

A Spectre alapvetően az arra érzékeny processzorgenerációkban használta ki a hyperthreading korlátait, miszerint a HT esetében a processzorgyártók nem minden regisztert dupláztak meg. Szerencsés együttállás esetében az egyik félprocesszoron futó alkalmazás beleláthat a másik félprocesszoron futó alkalmazás adatainak egy részébe.

Ez azonban rögtön két kihívást is támaszt a Spectre sérülékenységét kihasználni szándékozó támadó számára:

1. Gondoskodni kell arról, hogy a kártékony alkalmazás pont a megfelelő időben, pont a megfelelő támadandó applikációval egy időben ütemeződjön egy cloud-infrastruktúra megfelelő szerverének megfelelő processzorára. Ehhez először el kell érni, hogy a kártékony alkalmazás egyáltalán felkerüljön az adott infrastruktúrára, és utána el kell érni, hogy a hypervisor pont a megfelelő módon ütemezze azt. Nem lehetetlen, de igen komplex kihívás. Ráadásul a kezdetleges védelmi megoldás igen hamar megérkezett: tiltani kell a hyperthreading használatát például BIOS-beállításokban, vagy a kezdetleges hypervisor-javítócsomagok ugyanezt tették titkoltan. Nem véletlen, hogy az első, kezdetleges átmeneti javítások 20-40%-os teljesítménycsökkenést jósoltak, ez a pont annyi, mint amit a hyperthreading használata általában adna pluszban. A hypervisor patch gyakorlatilag letiltotta a hyperthreading használatát. A későbbi javítások természetesen sokkal kifinomultabbak lettek, biztosítva azt, hogy az ütemezés során egy processzor például két hyperthread felére csak azonos virtuális géphez tartozó processzek ütemeződjenek (Sibling Scheduler).

2. Ha valamilyen csodával határos együttállásnak köszönhetően mégis sikerül a Spectre segítségével részadatokat gyűjteni egy másik applikációból, azokból valami ténylegesen használható konzisztens információt generálni szintén komplex feladat, ami a megtámadott applikáció mély ismeretét igényli.

Mivel a Spectre estében már a sebezhetőség kihasználása is rettentően bonyolult, és a segítségével szerzett adatok használhatóvá tétele még bonyolultabb, nem csoda, hogy bár ezen sérülékenységek néhány éve nagy publicitást kaptak, általuk végrehajtott valóban sikeres támadásokról nem található információ a cikk írásának idején.

5.2. SS7 signalling és más protokoll-szintű támadások

A napjainkban használt telekom- és internetprotokollok jelentős része a 80-as évekből származik. Ugyan átestek különböző frissítéseken, új verzióik jeletek meg, de az alaplogika többnyire nem változott jelentősen az elmúlt évtizedekben. A kezdeti verziókba nem ritkán mai szemmel nézve csak korlátozottnak nevezhető védelmi mechanizmusok voltak beépítve, ugyanakkor ezek nagyon jól szeparált átviteli hálózatokon futottak (akkoriban sokszor nem is IP-transzporton). Később, amikor ezek a protokollok átkerültek IP-transzportra, vagy a már amúgy is IP-alapú protokoll publikus(abb) IP-hálózatra került, a transzport védelme IP-szinten ugyan megvédte a protokoll átvitel alatti integritását, de a protokoll logikájának alapvető sérülékenységét nem szüntette meg, a „poisoning” és hasonló módszerek továbbra is használatosak [10].

Erre egy példa a Syniverse feltörése [11]. Ez a cég többek között szöveges üzenetek továbbításával foglalkozik. A publikus információk között ugyan nem lehet arra konkrétumukat találni, hogy a megtámadott platformon a szöveges üzenetekhez milyen szinten férték hozzá, de mivel alapvetően az SMS is MAP-protokollra épül, gyanítható, hogy itt is valamilyen SS7/MAP-protokollgyengeséget használtak ki, vagy a megtámadott platformon esetleg a szöveges üzenetek továbbküldése alatt hozzáférhettek a tartalomhoz is. Az irodalomjegyzékben bővebb információ is található az SS7 protokollok támadási módszereiről [10].

Ez a példa arra is felhívja a figyelmet, hogy egy összetett támadás során lehetséges, hogy az infrastruktúra és az applikáció megtámadása is csak egy eszköz, a valódi cél egy a telekomot használó réteg támadása, mint például SMS-t használó banki multifaktoros hitelesítő rendszerek az SMS-ek lehallgatása vagy eltérítése által.

Egy másik példa a tavaly történt Facebook-leállás [12]. Ebben az esetben ugyan semmilyen jel vagy nyilatkozat nem utal arra, hogy ez valóban szándékos támadás lett volna, viszont a leállás oka szintén visszavezethető a BGP-protokoll gyengeségére. Az elérhető információk alapján ebben az esetben egy félrekonfigurálás történt, de a BGP-protokoll működéséből adódóan a helyreállás jelentős időt vett igénybe. Egy sikeres infrastruktúraszintű támadás második lépéseként a megfelelő jogosultságok megszerzése után ez egy remek szándékos támadási kísérlet eredménye is lehetett volna a rendelkezésre állás ellen.

6. OSS (Open Source Software) és egyéb „3rd party”-komponensek használata

Az előző fejezetek alapján látható, hogy a technológiai és üzletmenetbeli fejlődések hatására hogyan nőtt a telekom-infrastruktúrák sebezhetősége, ugyanakkor milyen komplex folyamat tud lenni egy támadás. Röviden azt is érintettük, hogy a sérülékenység elleni védelem már a teljes szoftverfejlesztési életciklust át kell járja. Ez a fejezet a szoftverfejlesztési trendekben rejlő biztonsági problémákra hívja fel a figyelmet.

Az IT irányába mutató konvergencia nem csak a HW- és infrastruktúra szintű SW-használatban figyelhető meg, hanem az applikáció fejlesztési módszerek területén is. Az egyik ilyen legfontosabb trend az OSS-komponensek elterjedt használata minden szinten.

Infrastruktúra esetében:

– Host OS.
– Cloud infra SW (hypervisor és/vagy container alapú virtualizációhoz).
– Néhány BIOS- és FW-jellegű komponensben. Ezekben ugyan nem olyan elterjedt az OSS-komponensnek használata, de ha mégis egy ilyen szintű HW-közeli modulban jelenik meg OSS által okozott sérülékenység, annak a következménye még súlyosabb.

Applikációs SW szinten: az applikációs réteg elvileg kizárólag az adott telekomfunkcióra kellene koncentráljon, de technikai okokból itt is szükség van olyan rétegekre, ahol az OSS- komponensek elterjedt használata jellemző:

– guest OS hypervisor alapú virtualizáció esetén,

– micro OS konténeralapú virtualizáció esetén,
– különböző, már az applikáció által direktben használt megoldásoknál (adatbázisok, webszerverek, kommunikációs modulok stb.).

Az OSS-komponensek használatának vitathatatlan előnyei vannak a szoftverfejlesztés során [13]. Az elterjedten használt funkciók, amik szinte minden applikációban felbukkannak (adatbázis-kezelés, webszerver, kommunikációs interfészek, AAA stb.) többnyire elérhetőek valamilyen OSS-komponens implementációban. Logikus lépés ezeket felhasználni, így a fejlesztés felgyorsul, olcsóbb lehet, konvergensebb lesz, és ami a legfontosabb, hogy a beszállító koncentrálhat az applikáció által képviselt tényleges funkciók implementálására. (A cikk nem foglalkozik az OSS-komponensek különböző felhasználási, jogi feltételeivel.)

Biztonsági szempontból azonban az OSS-komponensek használata jelentős kockázati forrást jelent, amivel a gyártónak és az üzemeltetőnek is foglalkoznia kell. Amennyiben bármelyik felhasznált OSS-komponens sérülékenynek bizonyul, az potenciálisan veszélyezteti az összes applikációt is, amelyekben az adott OSS-komponens adott verziója került beépítse.

Egy átlagos telekom-applikációban az OSS-komponensek százasával, nem ritkán ezresével vannak beépítve, ezért a legelső követelmény, hogy a gyártó (és később az üzemeltető is) tisztában legyen azzal, pontosan milyen OSS-komponenseket is épített be. Ez az egyszerűnek hangzó követelmény a valóságban nem mindig ilyen egyszerű, gondoljunk csak a különböző tranzitív függőségekre [14], vagy az adott gyártó saját komponensei közötti függőségekre. A felhasznált komponensek listázására különböző scannerek állnak rendelkezésre, amik vagy a szoftverfejlesztés során vagy akár a már kész termék (pl. cloud image vagy konténer) szinten képesek listázni az összes beépült OSS-komponensverziót. Ezeket elterjedten használják a beszállítók és az üzemeltetők is, az automatikus kiszállítási és integrálási folyamatoknak (CI-CD) [15] ezek az eszközök is már tipikusan szerves részei.

A fent említett eszközök nem csak listázni tudják a felhasznált OS-komponenseket, hanem azt is meg tudják nézni, hogy kapcsolódik-e az adott verzióhoz valamilyen addig ki nem javított sérülékenység. Ez az ellenőrzés tulajdonképpen egy adatbázis-alapú ellenőrzés, az adott OSS-komponens verzió alapján egy keresés például CVE-jellegű adatbázisban, de a különböző scannereket gyártó cégeknek saját adatbázisaik is vannak.

Fontos megjegyezni:

• Az adott termékre az adott OSS-komponensekre a scanner által talált sérülékenységek csak potenciális sérülékenységekként kezelendők (original score a CVSS alapján [16]). A valóságban ugyanazon OSS-komponens sérülékenységi hatása teljesen eltérő lehet attól függően, hogy az adott OSS-komponens milyen applikációba van beépítve, pontosan milyen funkcióit használják annak. (Expert score, a beszállító által végzett vizsgálat alapján, mivel egy sérülékenység vizsgálata során annak hatásláncát a valódi kontextusban szükséges elemezni). Tehát lehet például az automatikus integrációt blokkolni, amennyiben egy scanner valamilyen potenciális sérülékenyéget talál a vizsgált applikációban, ahova az OSS-komponens be van építve, de a valós fenyegetettség megállapításához mindig további vizsgálatok szükségesek.

• A tradícionális scannerek által végzett vizsgálat többnyire csak adatbázis-alapú ellenőrzést tartalmaz. Nem próbálják meg valóban leteszteltni vagy kihasználni az adott OSS-komponens sebezhetőségét. Gondoljunk csak bele, mennyire komplex feladat lenne például egy Spectre sebezhetőségi vizsgálata. Ehelyett egyszerűbb megnézni, hogy a felhasznált OSS-komponensverzió a CVE-adatbázis információi alapján már javítva lett-e, vagy még mindig sebezhető a Spectre által, vagy van e rá más nyitott sebezhetőség.

• Az OSS-komponensek keresése a scannerek által többnyire statikus, offline scannelés, nem igényli az applikáció futtatását, elég az image-eket scannelni. Nem összekeverendő a system hardening-el [17], amikor is egy futó alkalmazáson, futásidőben végezünk el bizonyos biztonsági ellenőrzéseket. A piacon elérhető scannerek nagy része kombinálja ezt a két funkciót: statikus analízis az applikáció image-re és dinamikus analízis a futó alkalmazásra.

Nézzünk egy konkrét példát, mi történik, amikor egy OSS-komponensre új sebezhetőséget jelentenek be. 2021 során az egyik ilyen legkomolyabb eset az év végén felfedezett log4j sérülékenység volt [18]. A nagy publicitásnak köszönhetően a bejelentés futótűzként terjedt el az üzemeltetői és beszállítói oldalon egyaránt. A beszállítók fel vannak készülve az ilyen esetekre és az új sérülékenységet a megfelelő processzeik alapján kezelik:

1. Még mielőtt egy új sérülékenység megjelenne, bármely időpillanatban szükséges egy friss belső adatbázis megléte, pontosan listázva, melyik termék milyen OSS-komponenseket használ. Emellett szükséges egy adatbázis a korábbi analízisek eredményeinek nyilvántartására (original score vs. expert score).
2. Amikor egy új sérülékenység megjelenik (pl. a lpg4j 2021 decemberében), le kell ellenőrizni, hogy a sérülékenynek ítélt OSS-komponens be van-e építve valamelyik termékbe.
3. Amennyiben a potenciálisan sérülékeny OSS-komponens jelen van, részletesebb elemzés szükséges, hogy a potenciális sérülékenység milyen hatással lehet a termékre, szükséges-e valamilyen javítás, kell-e frissíteni az OSS-komponens verzióját. Vizsgálni kell, hogy szükséges-e valamilyen átmeneti biztonsági intézkedés foganatosítása, amíg az új, javított OSS-komponensverzió nem áll rendelkezésre, illetve az frissítésre és kiszállításra kerül a termék egy javított verziójában, amennyiben a sérülékenyég valódi fenyegetések bizonyul a termékre nézve.
4. Dokumentálás, kiszállítás.

A példában említett log4j komponens egy nagyon széles körűen használt OSS-modul, ezért szinte minden te-

lekom-beszállító valamilyen applikációjában jelen van valamilyen szinten. Ez lehet direkt használat, de akár tranzitív függőség miatt is. A sérülékenynek minősítésnek köszönhetően minden beszállító rögtön megkezdte az analízist, a szükséges döntések nagy része még december folyamán megszületett (jelent-e valós veszélyt egy adott applikációra, ha igen szükséges-e új verzió kiszállítása, vagy van egyszerűbb megoldás is a sérülékenység elhárítására), ennek megfelelően amennyiben új verzió kiszállítása szükséges, az a cikk írásakor folyamatban van. A helyzetet bonyolította, hogy a log4j (2.x) ugyanazon sérülékenységéhez kapcsolódóan csak december folyamán négy CVE keletkezett, ami négy egymást követő log4j verziót eredményezett (2.15–2.17.1 a cikk írásának idején), tovább bonyolítva az integrálási és kiszállítási terveket.

Az OSS-sérülékenységek részletesen dokumentálva vannak (pl. CVE). Ez jó hír az elhárítás szempontjából, könnyebb eldönteni a fenyegetés mértékét a felhasználó applikáció szempontjából, könnyebb követni az OSS-hez kapcsolódó korrekciók állapotát, és úgy általában könnyebb belső folyamatokat építeni az új sérülékenységek kezelésére is.

Ugyanakkor ez a nyílt és rapid kommunikáció extra veszélyforrást is jelent, mivel – szintén a log4j példánál maradva –, a sérülékenységet kihasználók is részletes információkat kapnak nagyon gyorsan, a leírás alapján nem nehéz megfejteni, hogy milyen alkalmazásokban és milyen területeken használják a sérülékeny komponenst (log4j-t szinte mindenhol), és rögtön támadások indíthatók, amik megpróbálják kihasználni az átmeneti helyzetet, amíg a sérülékenységet nem foltozzák be, vagy átmeneti biztonsági intézkedéseket nem hoznak. Nagyon sok új sérülékenység esetén ráadásul még a korrekciók megjelenése előtt elérhetővé válnak automatizált eszközök a sérülékenység kihasználására, így különösebb előképzettség nélkül, nagyszámú rendszeren lehet próbálkozni, hátha valamelyikre nem jutott még el a szükséges adott korrekció.

Az OSS-komponensekkel kapcsolatos sérülékenységek kapcsán érdemes azt is elemezni, hogyan alakul ezek száma az applikációs SW életciklusa alatt (amibe az OSS-komponensnek be vannak építve).

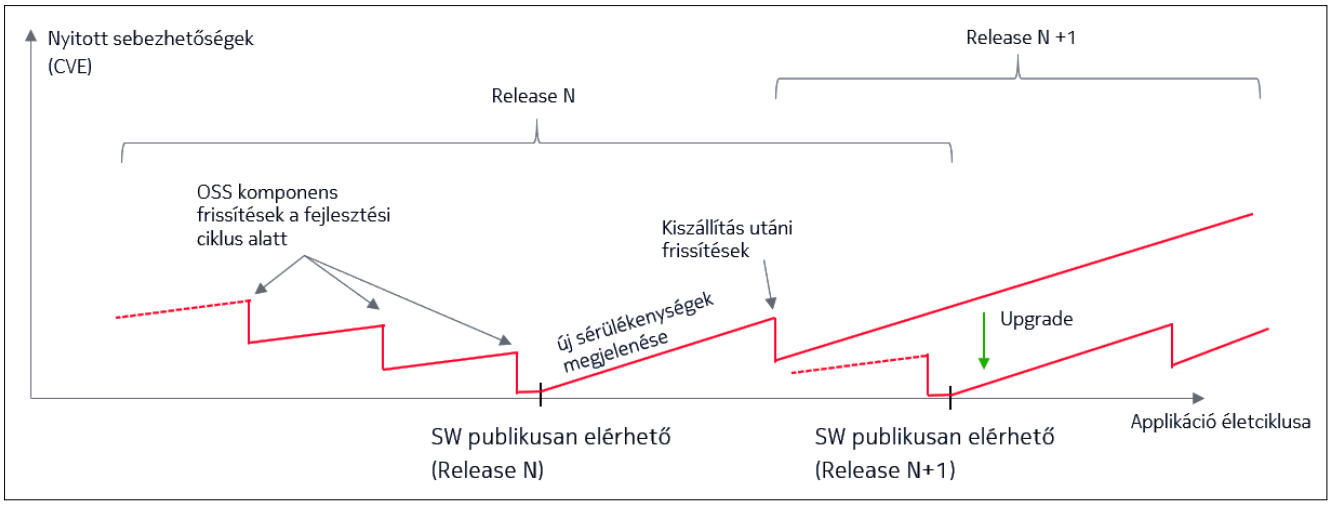
Mivel a SW életciklusa alatt a felhasznált OSS-komponensekre folyamatosan válnak ismertté új sebezhetőségek, azok száma a fejlesztési ciklus alatt és a kiszállítás után is folyamatosan növekszik. A *3. ábrán* ez a növekedés lineárisnak van feltüntetve, de a valóságban ez bármilyen jellegű növekedés lehet. A „nulla” sebezhetőséghez OSS-szempontból a legközelebb a publikusan elérhetőség kezdetén áll az applikáció. Ez a valóságban nem mindig szokott nulla lenni, különböző okok miatt, mint például:

– Lehetséges, hogy egy új CVE kapcsán még nem áll rendelkezésre a javított OSS-verzió. Ilyen esetben a beszállítónak el kell döntenie, tovább engedi-e a szoftvert, és majd később szállítja ki a korrekciót egy frissítés formájában, amint az OSS korrekció elérhetővé válik, vagy késlelteti az applikáció kiszállítását.
– A fentiekben szót ejtettünk arról, hogy a scannerek által talált sérülékenységeket potenciális sérülékenységként kell kezelni, és csak az applikációspecifikus analízis után dönthető el annak valós mértéke az applikációra nézve. Ezért a SW életciklusa alatt bármelyik időpillanatban előállhat az a helyzet, hogy a scannerek ugyan találnak valamilyen sérülékenységet az applikáció tisztán statikus, metaadatok felhasználása alapján történő átvizsgálása alapján, de a részletes analízis alapján azok a sérülékenységek nem jelentenek valós veszélyt, vagy a sérülékenységhez rendelt CVSS-pontszám [19] csökkenthető.

A nyitott OSS-sebezhetőségek száma az applikációban tipikusan az OSS-komponensek verzióinak frissítésével csökkenthető. A SW kiadása előtt ez a fejlesztési fázis szerves része, célul tűzve ki a kiadás pillanatában minimalizálni az applikációban lévő sérülékenységeket, a kiszállítás után pedig javítócsomagok formájában.

Amikor az adott applikáció új verziója elérhetővé válik, az természeténél fogva már kevesebb nyitott sérülékenységet fog tartalmazni a kiadás pillanatában, illetve

3. ábra OSS-sérülékenységek alakulása a SW életciklusa során.



a beszállítók is leginkább az aktuális vagy néhány legutolsó applikáció verzióra teszik elérhetővé a frissítéseket, ezért az OSS-hez kapcsolódó sérülékenységek minimalizálása céljából is érdemes az applikáció újabb verziójára váltani, általában a karbantartási szerződések is erre terelik a telekommunikációs applikációk üzemeltetőit.

OSS-felhasználási szempontból tehát a konklúzió az, hogy az egyértelmű előnyök mellett nem szabad elfeledkezni az általuk jelentett potenciális biztonsági fenyegetettségről sem, mivel manapság a telekom-termékeket érintő valós fenyegetettségek jelentős része valamilyen OSS-komponens által előidézett sérülékenységre vezethető vissza. (Nem számolva ide az alapvető biztonsággal kapcsolatos irányelvek be nem tartását a fejlesztés-kiszállítás és üzemeltetés során.)

Amennyiben egy beszállító OSS-komponensek felhasználása mellett dönt, elengedhetetlen az ehhez szükséges háttér folyamatos biztosítása is a biztonság megőrzése érdekében, mivel az OSS-t tartalmazó applikáció sebezhetetlensége folyamatosan inflálódik az életciklusa során. Ennek főbb követelményei:

– megbízható adatbázisok a felhasznált OSS-komponensekről és azokhoz tartozó korábbi elemzések eredményeiről,
– magas fokú automatizáció a gyors reagálás érdekében új sérülékenység megjelenése esetén,
– scannerek és más biztonsággal kapcsolatos eszközök integrációja a fejlesztési-kiszállítási láncba (CI-CD),
– megfelelő, transzparens kommunikáció a beszállító és az üzemeltetők között,
– szükséges erőforrások tervezése, a már kiszállított termékek teljes élettartama alatt a felhasznált OSS-szoftverkomponensekre jövőben felfedezett sérülékenységek foltozására.

A fentiekben ugyan az OSS-komponensek által jelentett kockázatokkal foglalkoztunk, de hasonló függőség fennállhat bármilyen, a beszállító által más gyártótól vásárolt szoftverkomponens esetében is. Ebben az esetben a jelentős különbség az SLA (Service Level Agreement) megléte, szemben az OSS-jellegű felhasználással.

7. Infrastruktúrán kívüli eszközök

Az előző szakaszok kizárólag telekom-infrastruktúrára koncentrálnak. A kezdetleges digitális rendszereknél a biztonsági kihívások nagy része valóban az infrastruktúrához kapcsolódott inkább, ide nem értve a terminál (mobiltelefon) fizikai megszerzését, vagy az abban tárolt adatok megszerzését valamilyen egyszerű klónozási módszerrel, amikor a fizikai eszköz megszerezhető. Ez főleg annak volt köszönhető, hogy ezek a végfelhasználói eszközök túlságosan buták voltak, üzemeltetési szempontból többnyire elég volt a köztük lévő szabványos interfészek megfelelő védelme.

Napjainkra azonban a végfelhasználói eszközök is hatalmas fejlődésen mentek keresztül, az IT-eszközök felé történő konvergencia itt is megfigyelhető, a gyártóspecifikus operációs rendszert használó kevésbé okos eszközök jelentős része a telefonok közül eltűnt, gyakorlatilag két fő opció maradt meg: Android és iOS.

Az infrastruktúrára leselkedő veszélyek nagy része ezekre is érvényesek, hiszen ezek is nagy mértékben használnak OSS-komponenseket, a nagyfokú hasonlóság miatt egy esetleges sérülékenység egyszerre túl nagy részét érintheti a használatos eszközöknek.

Az infrastruktúra beszállítója és a telekommunikációs hálózat beszállítója néhány kivételes esettől eltekintve (blokkolás, terminálmenedzsment-rendszerek használata) nem tehető felelőssé ezeknek az eszközöknek a biztonsági szempontból napra készen tartásáért, az nagy részben a felhasználó és az eszköz gyártójának a felelőssége.

Ezen cikk elsődlegesen ugyan a telekommunikációs hálózatokkal foglalkozik, de amikor ezen hálózatok sebezhetőségét tárgyaljuk, nem hagyhatjuk figyelmen kívül, hogy a telekom-hálózatok elleni támadások jelentős része a valóságban nem is a hálózatot támadja, hanem magát a végfelhasználói eszközt, gondoljunk csak a nagy figyelmet kapott NSO „zéro click”-megoldásaira az elmúlt évből [20]. Az ilyen és hasonló megoldásokkal kifejezetten a végfelhasználói eszközöket támadják, kihasználva azoknak a nagyfokú hasonlóságát és moduláris, sokszor OSS-komponenseket is tartalmazó felépítését.

Az 5G elterjedésével ez a probléma még több figyelmet kell kapjon. Az 5G egyik jelentős felhasználási területe az IoT-kommunikáció, ahol 5G-modemeket használnak. Tehát végfelhasználói eszközök alatt nem csak a tipikusan Android/iOS-alapú okostelefonokat értjük, hanem az 5G-modemeket is, amik a sima okostelefonokra leselkedő veszélyek mellett még több problémával szembesülnek. Biztonsági szempontból pont ezek a veszélyesebbek mert:

– nagyon sok típus létezik, sokszor aránylag kevesebb példányszámban gyártva a nagy mennyiségben gyártott mobiltelefonokhoz képest;
– gyártói oldalról rövidebb idejű támogatás (az alacsonyabb példányszámok vagy egyéb végbemenő folyamatok miatt, mint pl. portfoliótisztítás, merger és felvásárlások miatt);
– ha rendelkezésre is áll a biztonsági javításokat tartalmazó FW, ritkább frissítés nemtörődömség vagy technikai limitációk miatt (pl. távoli FW-frissítés nem mindig támogatott vagy a nem hivatalos FW-en futó eszközök).

Amikor biztonsági szempontból beszélünk elsődlegesen adatátviteli eszközökről, például modemek esetében, akkor érdemes felhasználni a Wi-Fi-modemekkel kapcsolatban már megtanultakat. Bár a rádiós technológia különböző, a kihívások hasonlóak. Éppen ezért ami kihívásokat a mostani Wi-Fi-eszközök támasztanak, azokkal lehet számolni az egyre nagyobb számban elterjedő 5G-modemek eseteben is, például az IoT-alkalmazásokban:

• „Don’t upgrade if not broken” – egy átlagos felhasználó milyen sűrűn frissíti például a mosógépében vagy

klímájában lévő modem SW-verzióját? Egy ipari környezetben elvárható, hogy egy autógyár gyártósorán a felhasznált adatátviteli eszközök az IoT-láncolatban naprakész biztonsági frissítéseket kapjanak, de egy átlagos felhasználó estében ez már nem biztosított, amíg valami konkrét hiba nem lép fel, általában nem szokták az ilyen eszközök SW-ét frissíteni.

• Mi a célja a támadásnak? (Ez igaz a telefonokra is, nem csak a modem-jellegű eszközökre). Az adott eszközön (pl. egy telefonon vagy ipari kamerán) megtalálható-átfutó információ megszerzése, vagy a megtámadott eszköz csak azért fontos, mert annak van közvetlenebb hozzáférése a valóban megtámadni kívánt hálózathoz? (Például egy színes LED-szalagot vezérlő szerkezet Wi-Fi-kontrollal önmagában nem különösen csábító célpont, de ha ez az eszköz közben hozzáfér egy belső hálózaton folyó kommunikációhoz, akkor már jó ötletnek tűnhet azt snifferként vagy poisoning célokra használni, vagyis a megtámadott eszköz nem a cél, csak egy belépési pont.)

A fentiek alapján a végfelhasználói eszközük sebezhetősége ugyan elsődlegesen a felhasználók problémájának tűnhet (és többnyire az is), de a valóságban a megtámadott végfelhasználói eszközök jelenléte a hálózati infrastruktúrára is veszélyt jelenthet (nem is beszélve a felhasználói elégedetlenségről, amikor a megtámadott végfelhasználói eszközök miatti problémákért a szolgáltatót próbálja a felhasználó felelőssé tenni).

A nem rendeltetésszerűen működő eszközök az infrastruktúra elleni támadásoknak az egyik kiindulópontja is lehetnek. Ilyen veszélyek lehetnek például nagy mennyiségű megtámadott telefon botnetté alakítása, DDOS-támadásra való felhasználása a hálózat ellen.

8. Összefoglalás

Az elmúlt évtizedek során a telekommunikációs infrastruktúrák hatalmas fejlődésen mentek keresztül. Miközben a tényleges applikációs logika (2-3-4-5G) egyre komplexebb lett, egyre több hálózati elem egyre többrétűbb egymáshoz kapcsolódását igényelve, a használt infrastruktúra a kezdetleges gyártóspecifikus HW- és SW-kombinációtól eljutott a teljesen IT-alapú komponensek széles körű felhasználásához. A vitathatatlan előnyök mellett, mint pl. költségoptimalizálás, fejlesztési sebesség, karbantarthatóság, a legújabb technológiák gyors bevezetése és az IT-konvergencia a lehetséges fenyegetések jellegét is megváltoztatta.

A telekom-applikációs logikát fenyegető veszélyek mellett már legalább ugyanolyan, vagy még annál is nagyobb mértékben kell foglalkozni az infrastruktúrát érintő veszélyforrásokkal. Ezeknek célja lehet maga a telekom-infrastruktúra (CIA-elemek bármelyike), valamely az adott telekom-infrastruktúrát használó telekommunikációs alkalmazás logikája, felhasználva az infrastruktúra gyengeségeit, vagy akár egy telekom-szolgáltatásokat használó, de teljesen más iparág részét képző szolgáltatás (például pénzügyi szektor, gyártásautomatizálás, egészségügy, fogyasztói eszközök stb.).

Hivatkozások

[1] SDL: *https://portal.etsi.org/Services/Centre-for-Testing-Interoperability/ETSI-Approach/Specification-Languages/SDL*
[2] Nokia DX 200: *https://en.wikipedia.org/wiki/Nokia_DX_200*
[3] ATCA: *https://en.wikipedia.org/wiki/Advanced_Telecommunications_Computing_Architecture*
[4] ETSI MANO: *https://www.etsi.org/technologies/open-source-mano*
[5] CVE-adatbázis: CVE-CVE (mitre.org)
[6] CIA-hármas: What is the CIA Triad and Why is it important? | Fortinet
[7] ETSI-szabványok: ETSI – Standards, mission, vision, direct member participation
[8] Spectre & Meltdown: *https://meltdownattack.com/*
[9] L1TF: *https://www.intel.com/content/www/us/en/architecture-and-technology/l1tf.html*
[10] SS7-támadások: A Step by Step Guide to SS7 Attacks – FirstPoint (firstpoint-mg.com)
[11] Syniverse story: Syniverse quietly admits it was hacked for five years | Light Reading
[12] Facebook story: *https://www.theverge.com/2021/10/4/22709260/what-is-bgp-border-gateway-protocol-explainer-internet-facebook-outage*
[13] OSS előnyei: *https://flosshub.org/sites/flosshub.org/files/Benefits%20and%20Drawbacks.pdf*
[14] Transitív függőségek: Transitive Dependency – an overview | ScienceDirect Topics
[15] CI/CD: What is CI/CD? Continuous Integration & Continuous Delivery in 2019 (katalon.com)
[16] CVSS: Common Vulnerability Scoring System SIG (first.org)
[17] Hardening: What is Systems Hardening? Read the Definition in our Security Glossary | BeyondTrust
[18] log4j: Log4j – Apache Log4j 2
[19] CVSS: *https://www.first.org/cvss*
[20] NSO zero click: Researchers call NSO zero-click iPhone exploit 'incredible and terrifying' | Engadget *https://www.news18.com/news/tech/explained-what-are-zero-click-hacks-and-why-are-they-such-a-menace-3988664.html*

Az ábrák elkészítéséhez felhasznált képek forrása:

ATCA HW illusztrálása: Security impact of the ATCA architecture adoption (webcast) – P1 Security

Gyártó specifikus HW illusztrálása: Nokia DX 200 | Core Network | Products | Carritech Telecommunications

IT HW illusztrálása: HPE ProLiant DL380 Gen10 server | HPE Store US

A szerzőről

CSORDÁS GÁBOR a BME Villamosmérnöki Karán diplomázott Irányítástechnikai és Robotinformatikai szakirányon. Másoddiplomáit bank- és pénzinformatika, majd IT-management területen szerezte. Fő szakterülete a beágyazott rendszerek mellett a telekommunikáció, több mint húsz éve ezzel foglalkozik. Részt vett az IP multimédiás alrendszer fejlesztésében és a telekom-applikációk cloud-platformra történő migrálásában. Később telekom-platformok és -infrastruktúrák tervezésével és méretezésével töltött közel tíz évet. Jelenleg a Nokia megoldásait érintő biztonsági kihívásokkal foglalkozik.

A mesterséges intelligencia és az új kódolási eljárások szerepe a videófeldolgozás fejlődésében

NELSON FRANCISCO, BORDÁS CSABA
MediaKind
Southampton, Egyesült Királyság | Budapest, Magyarország
{nelson.francisco; csaba.bordas}@mediakind.com

Kulcsszavak: videókódolás, mesterséges intelligencia, gépi tanulás, mozgásbecslés, MI-alapú tömörítési technológia

A videótömörítő rendszerek tervezésekor az a fő cél, hogy maximalizáljuk a videó minőségét egy adott bitráta esetében, vagy hogy a lehető legalacsonyabb bitráta mellett érjük el a célzott videóminőséget, mindezt jól meghatározott feldolgozási erőforrások felhasználása mellett. Mivel a gazdasági és környezetvédelmi szempontok gyakran szigorúan korlátozzák az erőforrásokat, a kódolótervezés hagyományosan kodekszakértőkre támaszkodott a heurisztikák és algoritmusok kifejlesztésében, hogy kiválasszák az egyes alkalmazások kódolási eszközeit az előre meghatározott hatékonysági vagy számítási paraméterek szerint. Ezek a heurisztikus módszerek a viszonylag alacsony tömörítési hatékonyságú kódolási eszközök egyszerű letiltásától kezdve a kódolási módok, hivatkozások, mozgásbecslés (Motion Estimation, ME), keresési tartományok vagy blokkméretek közvetlen korlátozásán át terjedhetnek, akár globálisan, akár helyi szinten, a forrás általános jellemzői alapján.
Bár ez a tradicionális megközelítés kiszámítható és következetes kódolási hatékonysági növekedést nyújt meghatározott tartalomtípus esetén, a méretezést megnehezíti annak megértése, hogy az egyes eszközök hogyan befolyásolják az egyes tartalomtípusok tömörítési hatékonyságát, és hogyan lépnek kölcsönhatásba egymással, különösen azért, mert a kódoló összetettsége és a tömörítési hatékonyság közötti kapcsolat nem lineáris.
A cikkbeli elemzés elvégzéséhez a MediaKind kihasználta a gépi tanulás (Machine Learning) lehetőségeit, olyan valós idejű MI-vezérelt kódolási döntéseket hozva ezáltal, amelyek sokkal hatékonyabbnak bizonyultak, mint bármely ember által meghatározott heurisztika vagy algoritmus.

1. Bevezetés

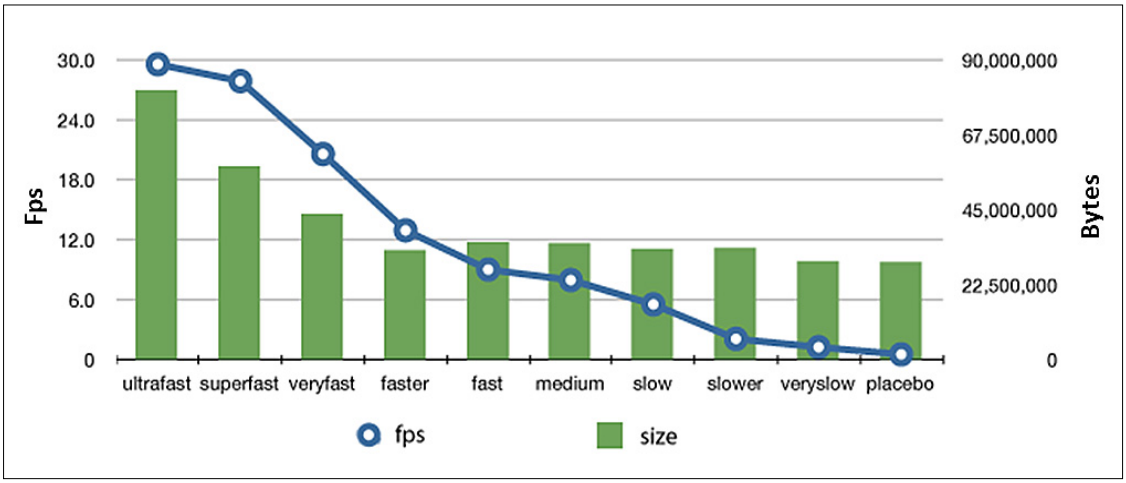
Minden évtizedben a videokodekek új generációja jelenik meg azzal a céllal, hogy elődjéhez képest megduplázza a tömörítési hatékonyságot. Az aktuális fejlesztéseket azonban főként az egyre számosabb és kifinomultabb kódolási eszközök kihasználásával hajtották végre, nem pedig a tömörítési paradigma teljes megváltoztatásával.

Az MPEG-2 [1] például nem használt intra-előrejelzést, amit a H.264 [2] esetében legfeljebb 9 különböző üzemmóddal [3] vezettek be. A HEVC [4] ezen módok számát 35-re emelte [5], a VVC [6] pedig tovább bővíti az intra-előrejelzés fogalmát azáltal, hogy az intra-előrejelzési módok maximális számát 87-re emeli [7]. Hasonlóképpen, az MPEG-2 és a H.264 is 16x16 pixeles makroblokkokat (MBs) [3] határozott meg, amelyeket azóta már felváltott a kódolási faegységek (Coding Tree Units, CTU) általánosabb koncepciója [5]. Ezek HEVC-ben akár 64x64 pixelt, VVC-ben pedig 128x128 képpontot is tartalmazhatnak [6]. Hasonló tendencia figyelhető meg a videokodek szinte minden aspektusában, a mozgásbecslés (Motion Estimation, ME) megnövekedett pontosságánál, a képcsoportstruktúrák (Group of Pictures, GOP) nagyobb rugalmasságánál, vagy akár a rendelkezésre álló hurokban lévő szűrők számában és összetettségében is.

Bár a feldolgozási teljesítményre vonatkozó követelmények növekedését részben teljesítették az állandó hardverfejlesztések, a felbontás és a képkockasebesség egyidejű növekedése azt hozza magával, hogy egyes alkalmazásokhoz nem elegendő kizárólag a hardverfejlesztésekre támaszkodni. A környezeti és gazdaságossági korlátok gyakran közvetlenül korlátozzák a rendelkezésre álló erőforrásokat és a működési költségeket, ami a kodek számítási összetettségének csökkentését igényli a videószállítási folyamat gazdaságilag életképessé tétele és szénlábnyomának minimalizálása érdekében.

Mindezek eredményeként a kódolófejlesztők gyakran szembesülnek azzal, hogy korlátozni kell a kódolási eszközöket annak érdekében, hogy megfeleljenek a célalkalmazás számítási erőforrás-költségvetésének [8,9]. Ez gyakorlatilag azt jelenti, hogy korlátozzuk az előrejelzési módok, referenciaképek vagy mozgáskompenzációs keresési tartományok számát, vagy egyszerűen letiltjuk a specifikáció által meghatározott kódolási eszközök némelyikét.

Fejlesztéseinkben ezért kihasználjuk a gépi tanulás (Machine Learning, ML) és az MI lehetőségeit, hogy optimálisan illesszük a használt kódolási eszközöket a rendelkezésre álló erőforrásokhoz, garantálva, hogy a feldolgozás az egyes alkalmazásokhoz mindig a leghatékonyabb legyen.



1. ábra Profilok összehasonlítása egy 60 mp-es HD-tartalomhoz, különböző x264-es készletekkel kódolva.

2. Tradicionális CODEC-konfigurálás

A videótömörítési mérnökök hagyományosan a tömörítési hatékonyság és a számítási összetettség hányadosának optimalizálása érdekében előre meghatározott paraméterkészletekre támaszkodtak [8,9]. A készletek olyan konfigurációs paramétereket határoznak meg, amelyek korlátozzák a rendelkezésre álló kódolási eszközöket, és amelyek mindegyike eltérő kompromisszumot kínál a számítási követelmény és a bitráta hatékonysága között. Ezt a megközelítést használja például az x264 [8] és az x265 [9], amelyek a H.264 [2] és a HEVC [4] videokodekek nyílt forráskódú implementációi. A beállításkészlet-opciók a placebótól (a legjobb minőség, nagyobb számítási összetettséggel) az ultragyors (alacsony komplexitás, korlátozott tömörítési teljesítmény) konfigurációig terjednek, lehetővé téve a felhasználó számára, hogy a kódoló hatékonyságát a rendelkezésre álló erőforrásoknak megfelelően állítsa be. Az, hogy melyik előre beállított beállítást használja, az alkalmazástól függ, mivel a nagyobb hatékonyság értékét gyakran kereskedelmi tényezők határozzák meg.

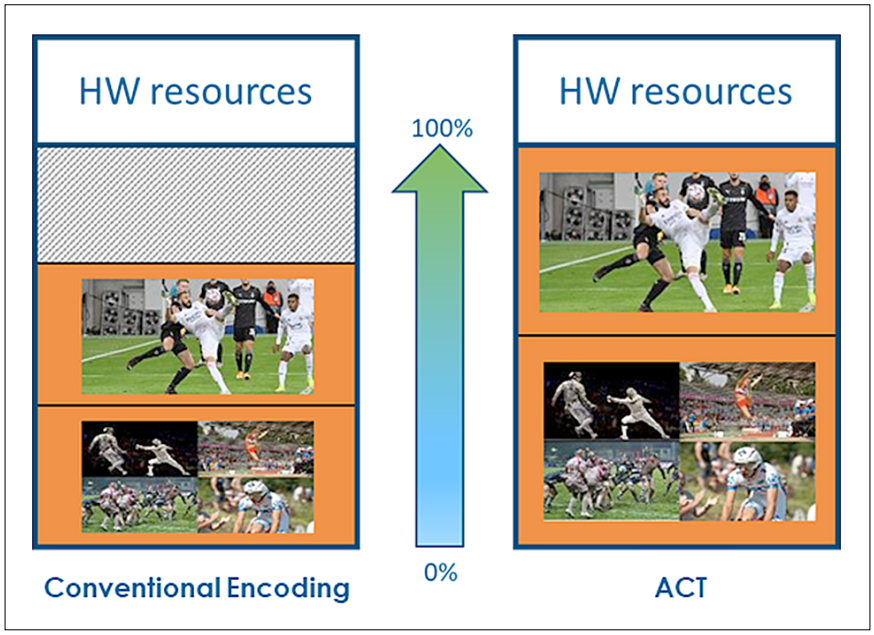
Az *1. ábra* az fps (frame per sec – képkocka/másodperc) számát és az ebből eredő tömörített fájlméretet mutatja egy 60 másodperc hosszú HD-tartalomhoz, amelyet az egyes rendelkezésre álló készletek segítségével kódoltak. Egyértelmű tendencia figyelhető meg mind a tömörítési hatékonyság, mind az erőforrás-követelmények tekintetében: az alacsonyabb profilok nagyobb tömörített fájlt eredményeznek (az alacsonyabb tömörítési hatékonyság eredményeként), de nagyobb számú képkockát tudnak feldolgozni másodpercenként ugyanabban az erőforrásban.

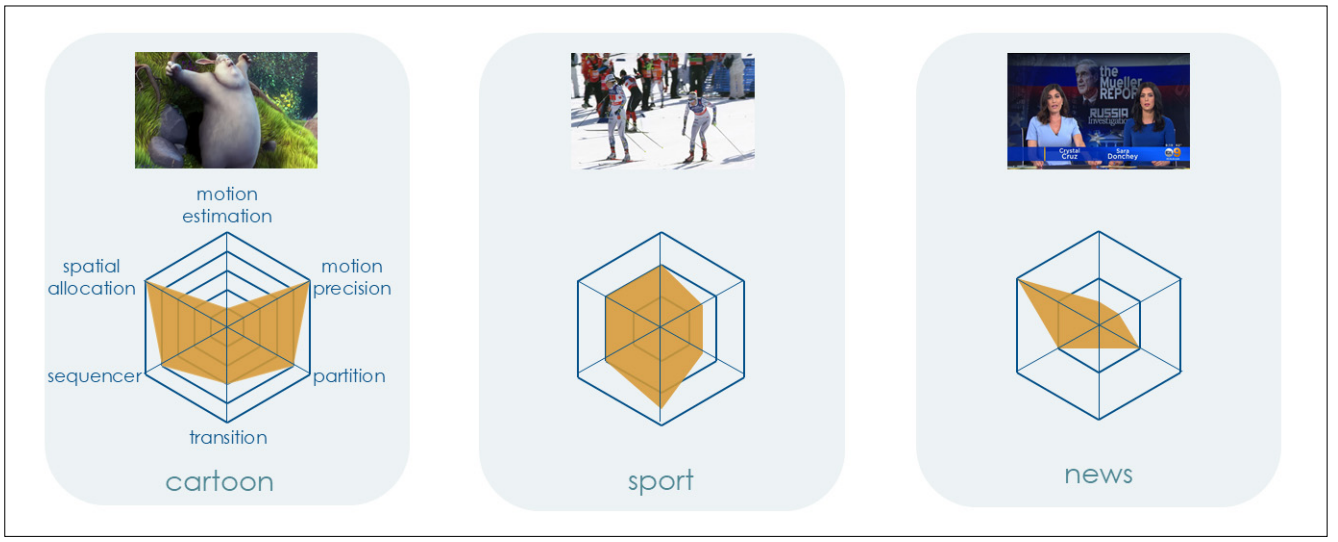
Ebben a példában azonban a „gyorsabb” („faster”) profil jobb tömörítési hatékonyságot ért el, mint néhány más készlet, amelyek több feldolgozást igényelnek. Ez rávilágít az előre beállított megközelítés egyik fő hiányosságára: az összetettebb készletek által hozzáadott extra eszközök és funkciók általában javítják a tartalom tömörítési hatékonyságát, de valójában káros hatással lehetnek bizonyos tartalmakra, a bemeneti forrás jellemzőitől függően.

3. ACT-MI-alapú tömörítési technológia (AI-based Compression Technology)

A konfigurációs, előre beállított megközelítés egyik problémája a kódoló számítási igényének beállításához nyújtott korlátozott lehetőségek. A legösszetettebb készlet, amely még mindig illeszkedik a rendelkezésre álló erőforrásokhoz, elméletileg a legjobb tömörítési hatékonyságot biztosítja, de előfordulhat, hogy az erőforrások jelentős részét nem használja, csak azért, mert a következő, jobb képminőséget nyújtó beállításkészlet csak kismértékben bár, de meghaladja a teljes feldolgozási költségvetést.

2. ábra Az ACT előnye több csatorna estén ugyanabban a rendszerben.





3. ábra Példa a különböző kódolási funkciók hatékonyságára a különböző tartalmakhoz.

Ez a probléma még nyilvánvalóbbá válik, ha több kódolót futtatunk ugyanazon a platformon, és nem működőképes különböző beállításkészletek üzemeltetésére különböző csatornákon. Az ACT-re vonatkozóan megállapított első követelmény tehát az volt, hogy a beállításokat granulárisabbá kell tenni, hogy maximalizáljuk az erőforrás-felhasználást és a tömörítési hatékonyságot *(2. ábra)*.

Az előző szakaszban említett előre beállított megközelítés másik hátránya, hogy az előre beállított készleteket átlagos, általános videóbemenetekre határozzák meg, és mind a komplexitás, mind a tömörítés hatékonysága jelentősen eltérhet a tartalomtípusok között. A különböző videóbemeneti jellemzők közvetlenül befolyásolják az egyes készletek által meghatározott kódolási eszközök hatékonysági előnyeit.

Amint azt a *3. ábra* mutatja, a rajzfilmek általában részletes textúrát és éles éleket mutatnak be, ami azt jelenti, hogy a pontos térbeli elosztás elengedhetetlen ahhoz, hogy pontosan ábrázolja a térbeli komplexitás variációit a képen. Továbbá, míg a mozgás általában könnyen nyomon követhető és megjósolható, tehát nem igényel túl sokat a mozgásbecsléstől, nagy mozgási pontosságra és nagyon pontos helymeghatározásra van szükség ahhoz, hogy elkerüljük az észrevehető kódolási hibákat.

Egy másik alkalmazást nézve, a térbeli kiosztás nem olyan fontos a sporttartalmakban, ahol a gyors kameramozgás, a gyors zoom és a térbeli és időbeli maszkolási hatások okozta elmosódás kevésbé észrevehető a nézők számára. Maga a mozgás azonban összetett és nehezen nyomon követhető az elzáródások, vágások és a világításváltozások miatt. Az olyan eszközök, mint a súlyozott előrejelzés, az extra erőforrások kiosztása a szekvenáló processzorba a leghatékonyabb GOP-struktúra meghatározásához, vagy több hivatkozás felvétele segíthet a mozgásbecslésben, és nagyon pozitív hatással lesz az általános tömörítési hatékonyságra.

A hír- és a stúdiótartalmaknál fontos a térbeli kiosztás, különösen átfedő grafikák esetén, de a mozgás könnyen nyomon követhető, tekintettel a korlátozott kameramozgásra és a jól meghatározott átmenetekre.

Annak érdekében, hogy jobban megértsük, hogyan befolyásolják a különböző tartalomtípusok az erőforrás-kiosztást, és határozzák meg a különböző típusú kódolási eszközök hatását a tömörítési hatékonyságban, az eszközöket kategóriákba csoportosították a közvetlenül érintett kodekterület szerint *(1. táblázat)*.

Az adaptív kvantálás és a sebességszabályozás például a bitráta-kiosztási stratégia részét képezi, míg a jelenetvágás észlelése és a GOP-tervező a szekvenálási kategóriába tartozik. Hasonlóképpen, míg a mozgásvektor finomítása a mozgáskompenzációs pontossági kategóriába tartozik, maga a mozgásbecslés, a referenciakezelés és a súlyozott előrejelzés a mozgásbecslési stratégia kategóriájába tartozik. Figyelembe kell vegyük, hogy függőség lehet a különböző kategóriákból származó eszközök között, mivel például a HEVC CU (Cod-

1.táblázat Kódolási eszközkategóriák.

Címke	A funkció leírása
ME	Mozgásbecslési stratégia
MCP	Mozgáskompenzálás pontossága
IPS	Predikción belüli keresés
PIPA	Kép particionálása
MD	Mód döntési stratégia
TQR	Átalakítás és kvantálás finomítása
BA	Bitráta-kiosztási stratégia
SEQ	Szekventálás
TM	Átmenet kezelése
PPE	Előfeldolgozás és becslés

4. ábra A vizsgált tartalmak.

ing Units) méreteinek korlátozása nemcsak a kép particionálási kategóriáját befolyásolja közvetlenül, hanem az átalakítás és a kvantálás finomítását is, mivel egyes átalakításméretek elérhetetlenné válnak. Az egyszerűség érdekében azonban elvonatkoztathatunk ezektől a függőségektől, anélkül, hogy ez veszélyeztetné az eredményeinket.

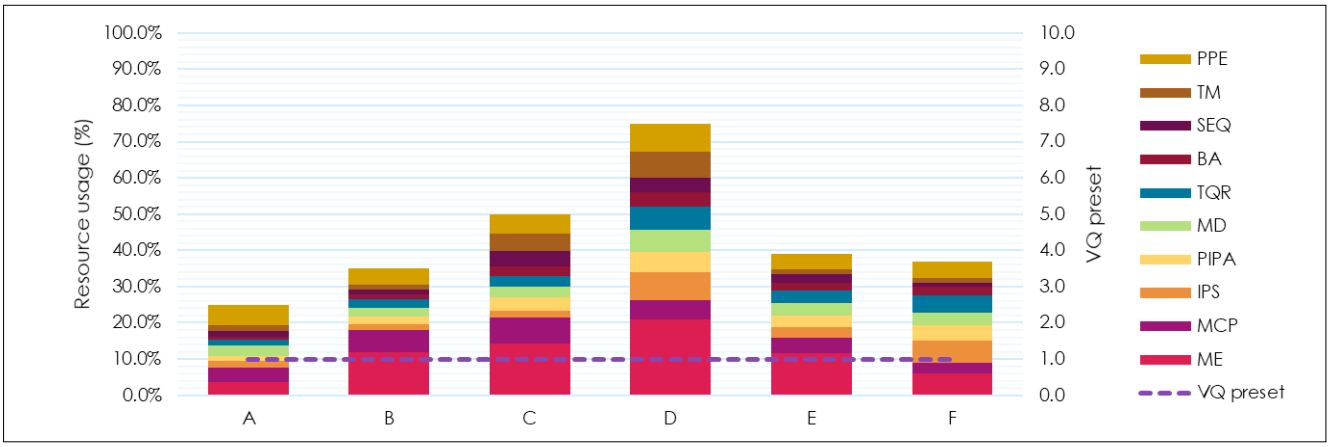
A *4. ábrán* összefoglalt, különböző térbeli és időbeli komplexitású tartalmak egy adott előre beállított videóminőségi értékre kódolásával az *5. ábra* egyértelműen mutatja az erőforrás-felhasználás és a bemeneti videó jellemzői közötti függőséget. Ebben a példában a *D* tartalomhoz közel háromszor több feldolgozásra volt szükség, mint az *A* tartalomhoz, amikor ugyanabban az előre beállított készletben kódolták őket. Emellett az egyes kódolási eszközkategóriák relatív erőforrás-felhasználása jelentősen változik a bemeneti videó jellemzőitől függően.

Az előre beállított megközelítés hatékonysági korlátai egyértelműen láthatóak az ábrán, mivel adott erőforrás-költségvetés mellett itt a *D* tartalom határozza meg a maximálisan használható beállításokat, ami elpazarolt erőforrást jelent bármely más tartalom esetén. Felmerül egy másik kérdés is: valóban a *D* tartalom a legrosszabb forgatókönyv? Ha a kódoló bemenetére a *D* tartalomnál is összetettebb videó kerül, átlépjük az erőforrás-költségvetést, ami katasztrofális következményekkel járhat, például át kell ugrani egy keretet, vagy összeomlik a kódoló. Ez gyakran azt jelenti, hogy az előre meghatározott beállítást még konzervatívabb módon kell kiválasztani, bizonyos biztonsági tartalék méretezésével, tovább növelve az erőforrás-pazarlást.

Az erőforrás-felhasználás figyelésével és a kódoló konfigurációjának dinamikus és valós idejű beállításával több eszköz engedélyezhető, amikor több erőforrás áll rendelkezésre. A gyakorlatban ez azt jelenti, hogy jobb videóminőségű előre beállított készletek használhatóak, ha a tartalom kevesebb erőforrást igényel. Ez a rendelkezésre álló feldolgozási erőforrások hatékonyabb felhasználását eredményezi, amint azt a *6. ábra* mutatja.

A korábbi megfigyelések alapján tehát meghatározzuk az optimalizálási problémát: azon kódolási eszközök és beállítások készletének megtalálása, amelyek maximalizálják a VQ-t (videóminőséget) egy adott tartalomhoz, egy adott t vagy $t+\Delta t$ pillanatban, miközben megfelelünk a valós idejű kódolás korlátozásának.

5. ábra Erőforrás-felhasználás eszközkategória szerint a különböző tartalomosztályokhoz.



Meghatározások:

• C a tartalomjellemzők vektora (tömörítési formátum, térbeli- és mozgáskomplexitások, gradienselemzés és időbeli variációs metrikák, online becsléssel a bemeneti forrás előzetes elemzése alapján).

• $p(C) \in \{p_1(C), .. , p_M(C)\}$ a kódolási előre beállított vagy kódolási paraméter-kombinációk listája a tartalomjellemzők függvényében, úgy, hogy:

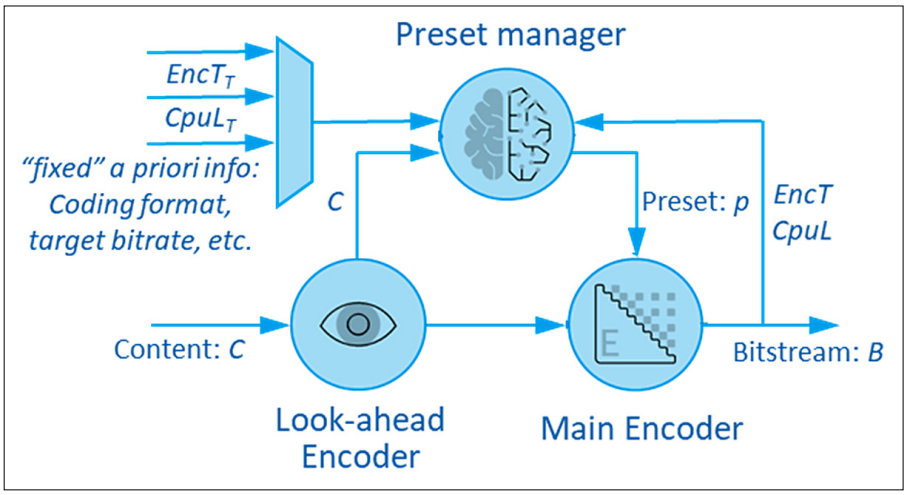
– A kódolási készlet a kódolási eszközszintek készlete/vektora; minden olyan kódolási eszközzel, amelyről feltételezzük, hogy legfeljebb *N*-szintű kódolási hatékonysággal/összetettségi kompromisszumokkal rendelkezik;

– A készletek listája a leggyorsabb/legalacsonyabb VQ-tól (1. készlet) a leglassabb/legjobb VQ-ig (előre beállított M) van meghatározva.

• *EncT* kódolási idő (ms/keret).

• *X*: kódolási futási célidő a kért kimeneti keret kódolási valós idejű küszöbértékének %-ában, $EncT_T = X * EncT_{realtime}$, $ahol\ 0 < X \leq 1$.

• *CpuL*: CPU/rendszer terhelésmérés.



7. ábra Az ACT-vezérlő blokkdiagramja.

Bármilyen *t* időpillanatban a C vektorjellemzőkkel rendelkező tartalom és egyéb a priori ismeretek, mint például a cél-bitráta, célhardver, R_T, stb. mellett meg akarjuk találni az optimális $p^*(C) \in \{p_1(C), .. , p_M(C)\}$ értékkészletet, amely maximalizálja a VQ-t az adott korlátok között: $EncT_T$ kódolási idő és $CpuL_T$ CPU terhelési cél:

$$\forall t, \begin{cases} p^*(C, \{HW, R_T, ..\}) = \max\limits_{p \in \{p_1 .. p_M\}} VQ\big(p(C, \{HW,\ R_T, ..\})\big) \\ EncT\big(p(C, \{HW, R_T, ..\})\big) = EncT_T \\ CpuL\big(p(C, \{HW, R_T, ..\})\big) \leq CpuL_T \end{cases}$$

Mivel a fő cél egyszerű módszer valós idejű futtatása volt, anélkül, hogy túl sok olyan erőforrást fogyasztanánk, amelyeket egyébként maga a kódoló használhatna, a C meghatározásánál a „lookahead”-kódolóban (a hagyományos kódoló prediktív része tipikusan 10-20 keret pufferelésével operál) a már kiszámított metrikákat használjuk a következőképpen:

• Intra-komplexitás vagy térbeli aktivitás:
$I = \sum_i^N \sum_j^M |x_{ij} - \mu|$, és $\mu = \frac{1}{NM} \sum_i^N \sum_j^M x_{ij}$.

• Inter-komplexitás:
$E = \sum_i^N \sum_j^M |x_{ij} - \hat{x}_{ij}|$,
ahol $\hat{x}_{ij}$ az x_{ij} mozgásbecsült blokk.

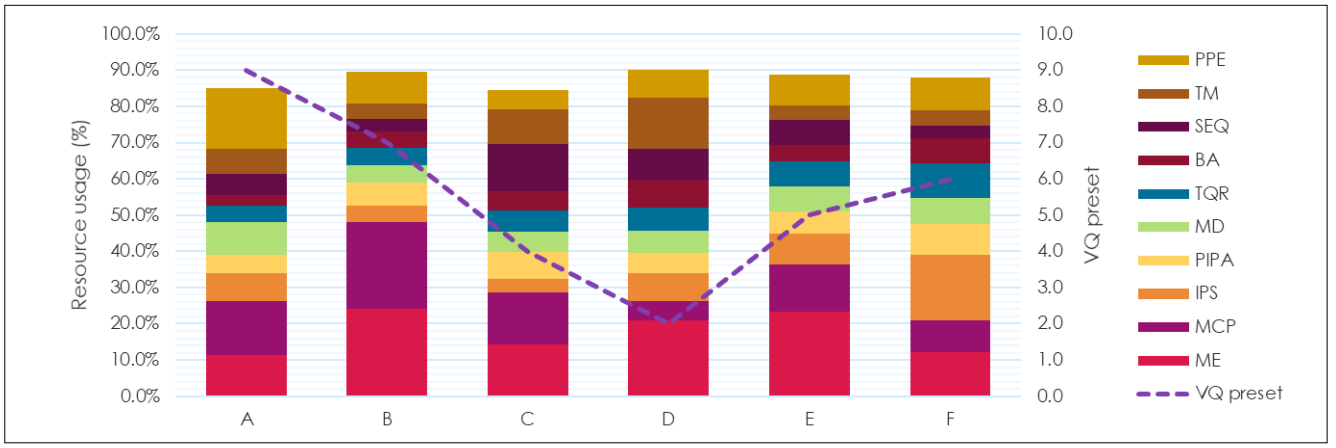
• Színátmenetek blokkolása:
$R = \sum_i^{N-1} \sum_j^{M-1} |x_{i+1,j+1} - x_{i,j}|$.

• Mozgásvektor költsége és entrópia.

Ezeknek a metrikáknak az a priori információkkal való kombinációja, például a kódolási formátum és a cél-bitráta lehetővé teszi a használandó konfigurációs paraméterek első becslését, a visszacsatolási hurok pedig korrekciós információkat nyújt arról, hogy a kódoló futásideje és a rendszer terhelése hogyan változik az idő múlásával, a további képkockánkénti korrekció végrehajtása érdekében, ahogyan azt a *7. ábra* mutatja.

Több kódolás futtatásával, a különböző kódolási eszközök konfigurációs paramétereinek a sokféle bemeneti tartalomhoz igazodó beállításával ezután neurális hálózat (Neural Network, NN) segítségével kinyerhető és meghatározható a „lookahead”-metrikák és az optimális

6. ábra Előre beállított készletek adaptációja tartalom szerint.



konfigurációk közötti korreláció. Az NN ezután felhasználható a bemeneti kép jellemzőinek megfelelő optimális konfiguráció valós idejű beállításainak végrehajtására is.

Az algoritmust mind a H.264, mind a HEVC esetében megvalósították és tesztelték. Ugyanazt a C tartalomjellemző-vektort használták, de a $p(C)$ konfigurációs paraméterek természetesen különböznek, mivel az egyes kodekeken elérhető eszközök nem azonosak.

A számítási erőforrások rögzítése mellett a javasolt módszer 18%, illetve 19%-os BD-SSIM (Bjøntegaard Delta Structural Similarity) nyereséget ért el a H.264 és a HEVC esetében, összehasonlítva egy rögzített előre beállított konfigurációval. A referenciaérték meghatározásakor a leghatékonyabb előre meghatározott beállítást használták, amely valós idejű kódolást garantál a rendelkezésre álló keretek között egy nagy, különböző tartalmakból álló tesztkészletben – míg az ACT minden képkockához beállította a kódolási paramétereket. Az eredményeket egyetlen csatornára számították ki, de amint azt korábban kifejtettük, a nyereségek még jelentősebbek lehetnek több csatornát futtató rendszerek esetében.

4. MI vezérelt HEVC CU felosztási döntések

Míg az MPEG-2 és H264 16x16 pixel méretű blokkegységeket fogadott el, amelyeket MBs-nek (Macroblocks) [3] neveznek, a HEVC új, általánosabb blokk-entitást vezetett be, amelyet CTU-nak (Coding Tree Unit, Kódolási Faegység) neveznek [8]. A CTU-k mérete 64x64, 32x32 vagy 16x16 pixel lehet, ami hierarchikusan tovább oszlik CU-kra (Coding Unit, kódolási egységek), az eredeti CTU mérettől 8x8 képpontig. Ez növeli a kódolás hatékonyságát azáltal, hogy lehetővé teszi az előrejelzések és a méretek jobb beállítását a tartalom helyi jellemzőihez, a számítási igények jelentős növekedésének rovására.

A valószínűleg használt blokkméretek pontosabb előrejelzése a kódolás előtt azt jelenti, hogy az összes lehetséges opciónak csak egy részhalmazát kell kiszámítania a fő kódolónak – hatalmas potenciális erőforrás-megtakarítással –, mivel ez a blokkméret-optimalizálási hurok jelentősen hozzájárul a kodek összetettségéhez.

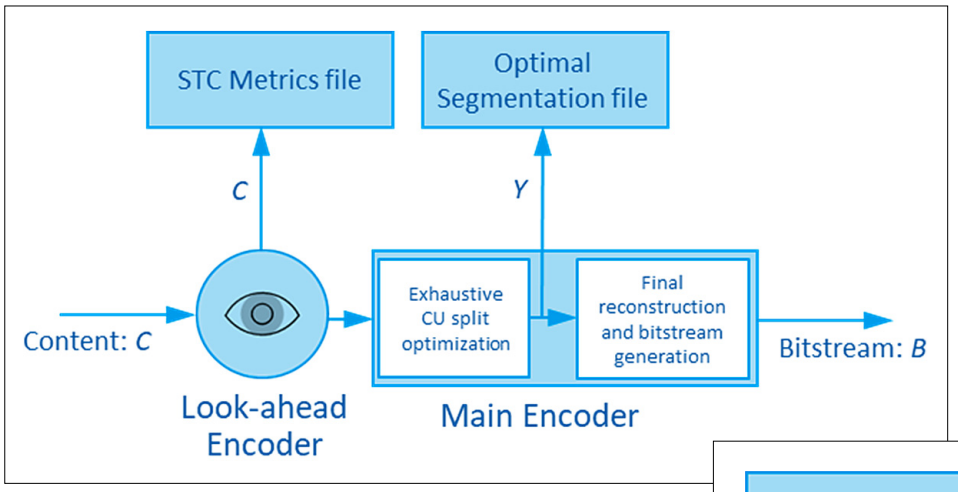
Nem meglepő tehát, hogy számos kutatás foglalkozott ezzel a problémával, és ez az a terület, ahol a gépi tanulás a legígéretesebb eredményeket hozta. Momcilovic és társai [11] olyan módszert javasoltak, amely akár 65%-os kódolási időcsökkenést ért el a kódolási ráta ingadozásának elhanyagolható növekedésével. Nem igényel betanítást, rendkívül adaptív algoritmust eredményez, amely dinamikusan reagál a videóbemenet változásaira. Egy másik tanulmány szerzői gépi tanuláson alapuló módszert javasoltak olyan funkciók felhasználásával, amelyek leírják a CU-statisztikákat és a képernyőtartalom al-CU-homogenitását [12]. A szerzők átlagosan 36,8%-os komplexitáscsökkenést értek el, mindössze 3,0%-os bitráta-növekedéssel. Chen és társai [13] egy gyors kódolási egységet (CU) javasoltak a HEVC-n belüli kódoláshoz egy mesterséges neurális hálózat (ANN) és egy támogató vektorgép (SVM) felhasználásával. Módszerükben a gépi tanulás szisztematikus megközelítést biztosított a korai CU-megosztás vagy -megszüntetés gyors algoritmusának kifejlesztéséhez a kódoláson belüli számítási komplexitás csökkentése érdekében. A Convolutional-Neural-Network (CNN) használata talán a legsikeresebb megközelítés volt. A szerzők [14]-ben egy kontrasztnyereség-ellenőrzési modellt javasolnak, amely megpróbálja megragadni a felismerhető struktúrák hatását a torzítás láthatóságára. Liu és társai [15,16] egy bonyolult neurális hálózaton alapuló gyors algoritmust javasoltak, hogy minden CTU-ban legalább kettővel csökkentsük a CU-partíciósmódok számát. Ez a módszer teljes sebességtorzulás-optimalizálást (RDO) használ. Az algoritmus az intra-kódolási idő 63%-át takarította meg átlagosan 2,66%-os BD-BR (Bjøntegaard Delta Bit-Rate) növekedési költségen. Javasolták a CNN-ek használatát [17] is, hogy megjósolják a 32x32 pixeles CTA-k 90%-os pontossággal történő felosztását.

A CNN-ek előnye, hogy lehetővé teszik a modell számára, hogy hatékonyan kinyerje a leghasznosabb funkciókat a bemeneti adatokból, ami akkor fontos, ha a hangsúly a legjobb pontosság elérésén van. A CNN-ek azonban számítási szempontból költségesek, és a GPU-k használatát igénylik a valós idejű kódoláshoz alkalmas átviteli sebesség eléréséhez. Ez azt jelenti, hogy nem biztosítanak mérhető teljesítménynövekedést a valós idejű kódoláskor, mivel a kódoló összetettsége már viszonylag alacsony [17].

A javasolt megközelítés középpontjában a modell pontossága és összetettsége közötti egyensúly megteremtése állt, tekintettel a valós idejű alkalmazásokban kitűzött célra. Így ahelyett, hogy CNN-eket használnánk a CU felosztásának megállapításához használt videójellemzők kinyerésére, újra felhasználtuk a „lookahead”-kódolóban már elérhető metrikákat. Ez a megközelítés korlátozhatja az előrejelzések pontosságát, de nem igényli a GPU használatát, így a módszer valóban általános és platformfüggetlen. Ugyanazokat a „lookahead”-metrikákat használjuk, mint az ACT-ben, nevezetesen a térbeli aktivitást, az inter-komplexitást, és az egyes 16x16 blokkokhoz kiszámított gradienseket, egy alul-mintavételezett képen. A CU QP (Compute Unit Quantization Parameter) és a kerettípus is szerepel a következtetéshez használt adatokban.

Különböző hálózati architektúrákat vizsgáltunk, különböző számú és kiterjedésű rétegekkel. A következtetés pontossága és a költségek közötti kompromisszumként egy 4 teljesen összekapcsolt réteggel rendelkező NN-t használtunk. Az első 3 rétegnek 50 neuronja van, amelyek ReLU (Rectified Linear Unit) aktiválási funkciót használnak, míg az utolsó rétegnek 21 neuronja van, és ez sigmoid aktiválási funkciót használ. Ez a hálózat átlagosan több mint 92%-os pontosságot ért el a CU-k 64x64 és 16x16 közötti felosztása esetén (a 8x8 nem osztható tovább).

Az NN-betanítást offline, GPU-val végezték egy több mint kétezer videósorozatból álló reprezentatív képzési



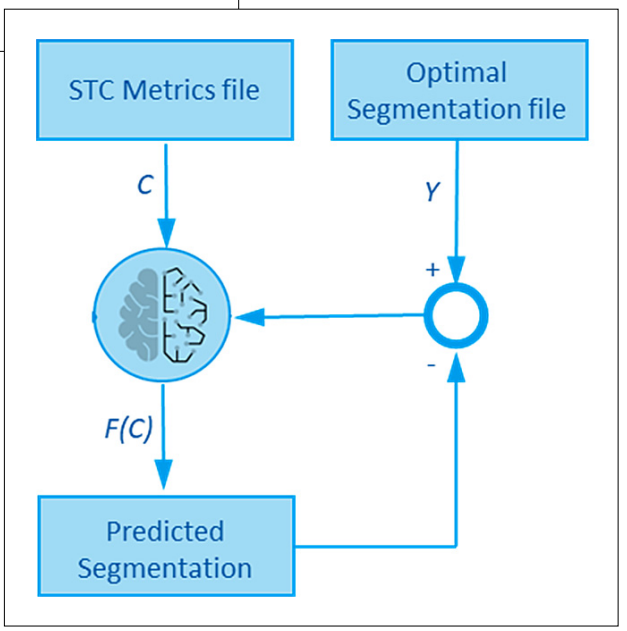
8. ábra
NN adatkészlet-generálás.

készleten. A bemeneti videót több QP-ponton kódolták, amelyek a teljes QP-tartományt lefedik és teljes RDO-keresést használnak, ahol az összes CU-méret engedélyezve van. Mind a lookahead-metrikák, mind az egyes CTU-khoz származó optimális szegmentációs jelzők naplózva voltak fájlokba mentve, amint azt a *8. ábra* mutatja.

Felügyelt betanítási megközelítést alkalmaztak, a *C* mérőszámokat az NN-be táplálták, hogy *F(C)* következtetést hajtsanak végre a CTU felosztásához. A kódolóból a teljes RDO kereséssel nyert optimális szegmentációt alapvetésként használják, amely lehetővé teszi a következtetett és az alapfelvetés közötti bináris keresztentrópia-vesztés meghatározását:

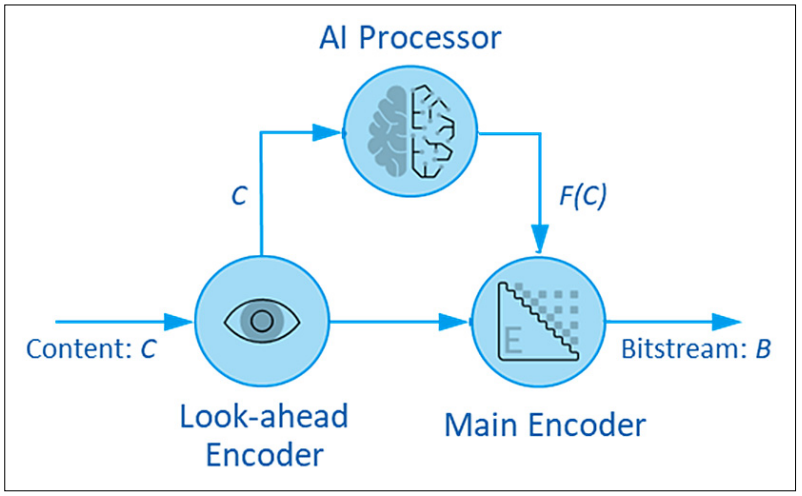
$$L = -\sum_{n} w_n \left[y_n \log \sigma(F(c_n)) + (1 - y_n)\left(1 - \log \sigma\left(F(c_n)\right)\right)\right]$$

Ez a megközelítés feltételezi, hogy minél közelebb van az NN döntése a teljes optimalizálási eredményekhez, annál optimálisabb lesz a megoldás. A súlygradiensek visszaemelése lehetővé teszi a sztochasztikus gradiens süllyedési optimalizálását pillanattal [18] az NN-súlyok gradiensei felett, interaktív módon csökkentve a teljes veszteséget. A *9. ábrán* felvázoltuk a betanítási folyamatot. A betanítási adatokat képzési és érvényesítési készletekre osztották, az adatpontok 80%-ával, illetve 20%-ával.



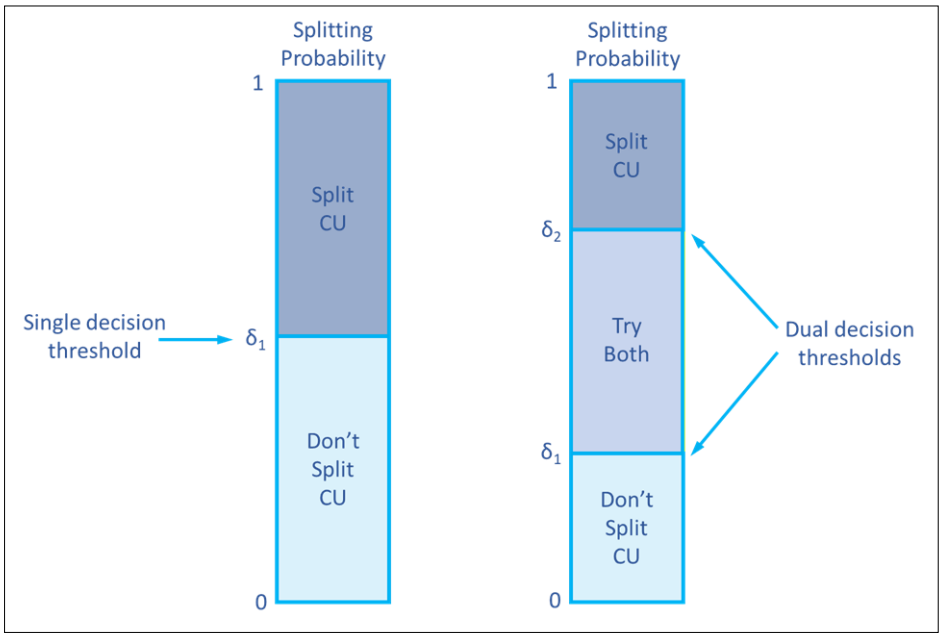
9. ábra
NN-betanítási folyamat.

10. ábra NN-következtetés a valós idejű kódolóban.



Miután az offline betanítás befejeződött, az NN-súlyokat a kódolóba táplálták, hogy a döntés valós időben elvégezhető legyen, egyszerűen a „lookahead”-kódoló adatainak felhasználásával, lásd a *10. ábrát*. Mivel NN viszonylag kicsi, többnyire lineáris rétegekre támaszkodik, amelyek hatékonyan megvalósíthatók SIMD-sel, azt jelenti, hogy a döntés nagyon alacsony költséggel elvégezhető a CPU-ban. Fontos, hogy ezzel a megközelítéssel a modell könnyen frissíthető minden alkalommal, amikor az offline betanítás olyan súlyokat eredményez, amelyek jobb pontosságot biztosíthatnak.

A módszerben alkalmazott valószínűségi megközelítésnek az az előnye, hogy skálázhatóvá teszi azt. Tulajdonképpen minden CTU esetében a következtetés határozza meg annak valószínűségét, hogy az egyes CU-kat szegmentálják-e vagy sem. A nullához közeli *F(C)* valószínűségek azt jelentik, hogy a CU szegmentálásának K+F költsége nem valószínű, hogy alacsonyabb lesz, mint a nem szegmentálás költsége, míg az 1-hez közeli valószínűségek azt jelzik, hogy a CU szegmentálása valószínűleg alacsonyabb RD-költséget eredményez.



11. ábra Egyedüli és kettős döntési küszöbértékek

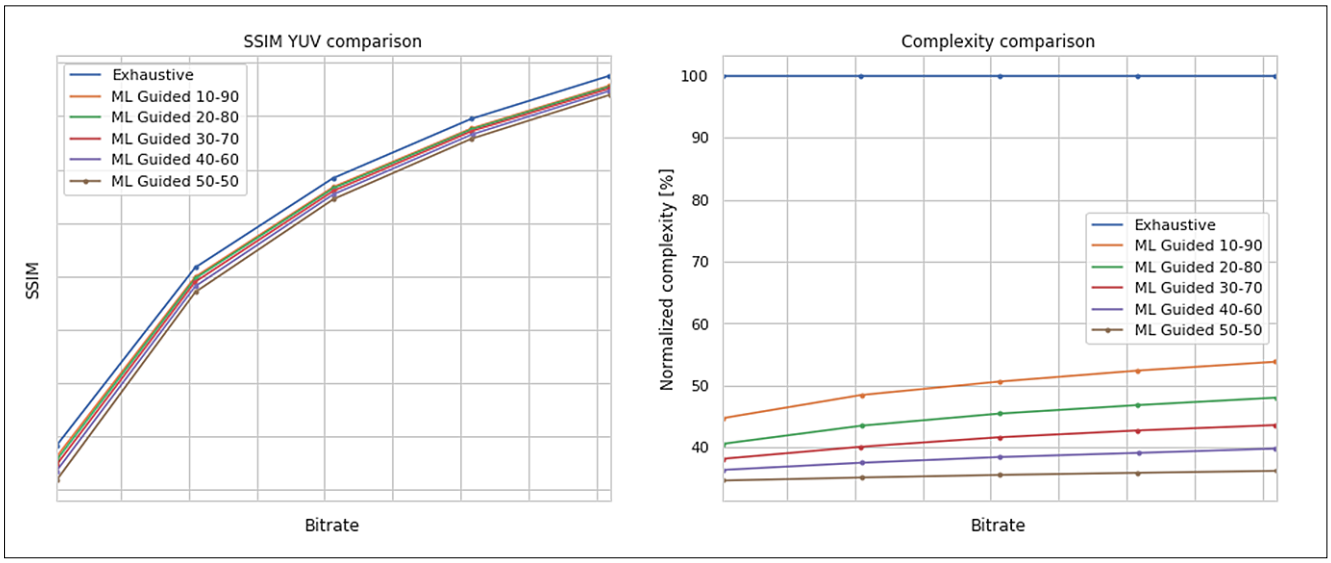
A legnagyobb számítási megtakarítást egyetlen δ_1 küszöbérték meghatározásakor érjük el. A CU-t akkor osztjuk fel, ha az $F(C) > \delta_1$, és nem osztjuk meg, ha az $F(C) < \delta_1$, tehát ebben az esetben minden CU-ra csak az egyik felosztási lehetőséget kell kiszámítani. A 0,5-höz közeli valószínűségek azonban nagy bizonytalanságot jelentenek, amely könnyen egy nem optimális döntéshez vezethet, ami ezek után befolyásolhatja a kódoló tömörítési hatékonyságát. Ez a probléma megkerülhető két különböző δ_1 és δ_2 küszöbérték meghatározásával, ahol $\delta_1 < F(C) < \delta_2$ esetén mind a felosztási, mind a nem felosztási lehetőségeket tesztelve viszonylag alacsony számítási költséggel csökkenthető az optimálistól eltérő döntés kényszerítésének valószínűsége *(11. ábra)*.

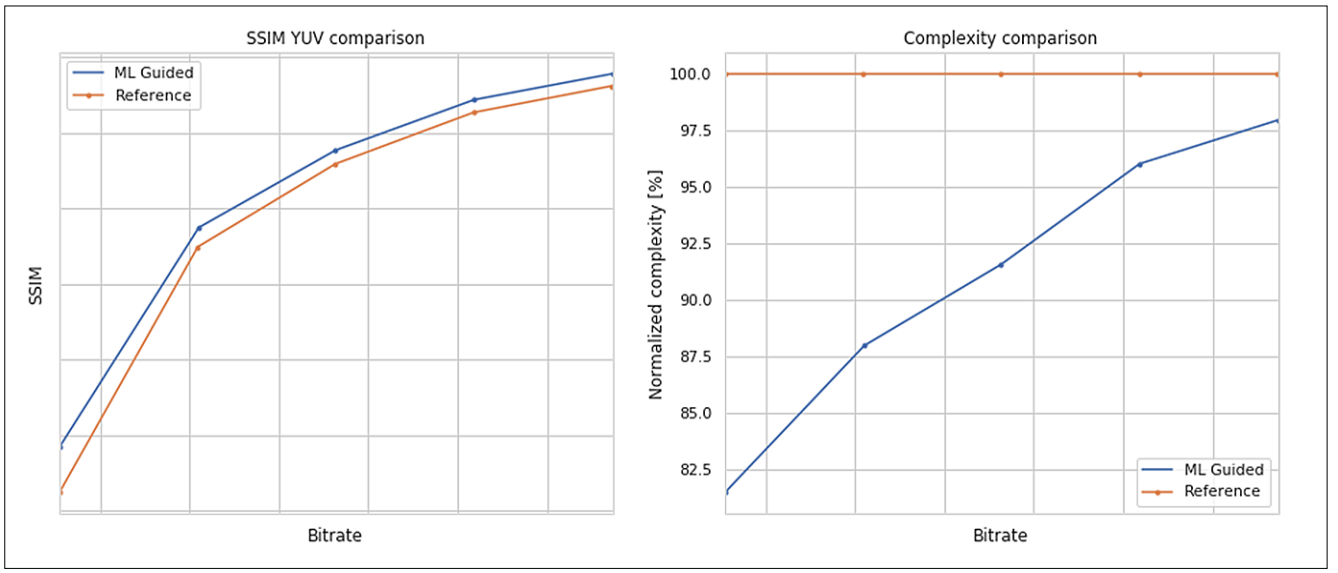
Az intervallum méretének növelése tömörítési hatékonyságot növelő döntésekhez vezet (kevesebb kockázatot vállalva több mód tesztelésével), míg az intervallum szűkítésével a döntéseket a számítási megtakarítások felé toljuk (lehetővé téve a kódoló számára, hogy több kockázatot vállaljon és több módot dobjon el). A rendelkezésre álló erőforrások visszacsatolási hurkának megtartásával ez az intervallum valós időben, az ACT-vel párhuzamosan módosítható.

Az alsó,*12. ábrán* összehasonlítjuk a képminőséget és a kódoló összetettségét, különböző valószínűségi küszöbértékek használata esetén. A grafikonokon látható eredmények több 3840x2160@60fps szekvenciával rendelkező tesztkészletek átlagát mutatják, különböző típusú tartalmak esetén, a sporttól a filmekig.

A képminőséget az SSIM-mel (Structural Similarity Index Measure) mérik. Hogy elvonatkoztassunk a használt kódolási platformtól, a kódolás összetettségét a kimerítő keresés (Exhaustive Search) összetettsége alapján normalizáltuk. Ez azt jelenti, hogy a 100% ugyanolyan összetettségnek felel meg, mint a teljes keresés, míg az 50% azt jelenti, hogy az új kódoló a teljes kereséshez szükséges számítás felét igényli (a feldolgozási erőforrások feleakkorák ugyanazon keretsebesség eléréséhez, vagy ugyanazon a hardverplatformon feleződik a képkockaidő). Az ML-vezérelt 50-50 görbe esetén $\delta_1=\delta_2=0,5$, míg az ML-vezérelt 10-90 esetén $\delta_1=0,1$ és $\delta_2=0,9$ értékeknek felelnek meg. Az összes konfigurációs paraméter változatlan marad, és az ML vezérelt konfigurációk összetettsége magában foglalja az NN következtetését és a kódolást, a referenciaértékkel való közvetlen és korrekt összehasonlításhoz.

12. ábra Tömörítési teljesítmény és kódoló komplexitás összehasonlítása különböző valószínűségi küszöbértékekkel.





13. ábra Videominőségi és számítási összetettségi eredmények egy UHD 60 fps tesztkészlethez.

Ahogy az várható volt, a legjobb videóminőséget a teljes kereséssel érik el. Agresszívebb ML-vezérelt 50-50 esetén akár 3 csatornát is kódolhatunk ugyanazzal a számítási kapacitással, míg a BD-SSIM csökkenés mindössze 5,5% körül van. A bizonytalansági intervallum növelése lehetővé teszi a tömörítési hatékonysági büntetés csökkentését, miközben továbbra is 50% feletti átlagos normalizált összetettségi csökkenést érünk el. Ez azt jelenti, hogy legalább 2 csatorna futtatható ugyanazon a hardverplatformon, és csak szerény 2%-kal csökken a BD-SSIM.

A *13. ábra* egy valós idejű alkalmazás eredményeit mutatja be, ugyanazzal a 3840x2160@60fps tesztkészlettel. A nagy felbontás és keretsebesség komoly hardvererőforrás-igényeket támasztanak valós idejű tömörítés eléréséhez, és még a nagy kapacitású, drága szervereken is csökkenteni kell a kodek eszközkészletét egy tipikus 1080p konfigurációkhoz képest. Az ML-vezérelt CU-felosztás által nyújtott számítási megtakarítások újra felhasználhatók más, korábban letiltott kódolási eszközök engedélyezéséhez, ami több mint 10%-kal javítja a kódoló tömörítési hatékonyságát, miközben a teljes számítási igények alacsonyabb maradnak, mint az eredeti konfigurációban. Vegyük észre, hogy a számítási megtakarítások jelentősebbek alacsonyabb bitráta esetén, mivel az alacsonyabb bitráta általában nagyobb CUs méreteket eredményez, lehetővé téve, hogy ugyanazon bizonytalansági intervallumban több felosztási lehetőséget ugorjunk át.

5. Összefoglalás

Ebben a tanulmányban gépi tanulási ML-technikák használatát javasoltuk a tömörítési hatékonyság és számítási komplexitás hányados javítására a valós idejű alkalmazásokhoz használt hagyományos videokódolókon.

A javasolt algoritmusok lehetővé tették a hatékonyabb feldolgozási erőforrások felhasználását a kódolási lehetőségek dinamikus szűkítésével és a kodek paraméterek módosításával. Egyes esetekben akár 50%-os számítási kapacitás megtakarítást mérhetünk hasonló tömörítési arányoknál, vagy akár 20%-os tömörítési hatékonyságnövekedést egyenértékű számítási követelmények esetében.

A kodek más területei, például az Intra/Inter/Skip döntések is használhatnák egy ilyen megközelítés előnyeit, a jövőben ezeket is meg szeretnénk vizsgálni.

Hivatkozások

[1] H.262: Information technology – Generic coding of moving pictures and associated audio information: Systems. www.itu.int. ITU. 2021-06.
[2] H.264: Advanced video coding for generic audiovisual services. www.itu.int. ITU. 2021-08.
[3] T. Wiegand, G.J. Sullivan, G. Bjontegaard and A. Luthra, „Overview of the H.264/AVC video coding standard," in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, no.7, pp.560–576, July 2003, doi: 10.1109/TCSVT.2003.815165.
[4] H.265: High efficiency video coding. www.itu.int. ITU. 2021-08.
[5] G.J. Sullivan, J. Ohm, W. Han and T. Wiegand, „Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, no.12, pp.1649–1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.
[6] H.266: Versatile Video Coding. www.itu.int. ITU. 2020-08.
[7] J. Pfaff et al., „Intra Prediction and Mode Coding in VVC", in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 31, no.10, pp.3834–3847, Oct. 2021, doi: 10.1109/TCSVT.2021.3072430.
[8] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, „Block Partitioning Structure in the HEVC Standard," in IEEE Transactions on Circuits and Systems for Video Technology, pp. 1697–1706, 2012.

[9] Loren Merritt et al,
„X264: A High Performance H.264/AVC Encoder”, 2006.
[10] D. Silveira, M. Porto and S. Bampi,
„Performance and energy consumption analysis of the X265 video encoder,” 25th European Signal Processing Conference (EUSIPCO), 2017, pp.1519–1523,
doi: 10.23919/EUSIPCO.2017.8081463.
[11] S. Momcilovic, N. Roma, L. Sousa, and I. Milentijevic,
„Run-Time Machine Learning for HEVC/H.265 Fast Partitioning Decision,”
in IEEE International Symposium on Multimedia, 2015.
[12] F. Duanmu, Z. Ma, and Y. Wang,
„Fast CU partition decision using machine learning for screen content compression,”
in IEEE International Conference on Image Processing (ICIP 2015), 2015, pp.4972–4976,
doi: 10.1109/ICIP.2015.7351753.
[13] Z.-Y. Chen, J.-T. Fang, Y.-C. Liu, and P.-C. Chang,
„Machine Learning-based Fast Intra Coding Unit Depth Decision for High Efficiency Video Coding,”
Journal of Information Science and Engineering, 2016.
[14] M.M. Alam, T. Nguyen, M. Hagan, and D. Chandler,
„A perceptual quantization strategy for HEVC based on a convolutional neural network trained on natural images”
2015, p.959918,
doi: 10.1117/12.2188913.
[15] Z. Liu, X. Yu, S. Chen, and D. Wang,
„CNN oriented fast HEVC intra CU mode decision,”
in IEEE International Symposium on Circuits and Systems, 2016.
[16] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang,
„CU Partition Mode Decision for HEVC Hardwired Intra Encoder Using Convolution Neural Network,”
in IEEE Transactions on Image Processing, 2016.
[17] J. Wenchan, Y. Ming, X. Ying, and L. Zhigang,
A Machine Learning Method for Optimizing Partition of Prediction Block in Coding Unit in H.265/HEVC, (2020).
doi: 10.4108/eai.27-8-2020.2297985.
[18] Kingma, Diederik and Ba, Jimmy,
Adam: A Method for Stochastic Optimization, (2020).
International Conference on Learning Representations.

A szerzőkről

NELSON FRANCISCO a MediaKind kutató mérnöke 2008-ban szerzett elektronikai mérnöki mesterdiplomát a Tras-os-Montes e Alto Douro Egyetemen (Portugália) és 2012-ben PhD fokozatot jelfeldolgozás témakörben a Rio de Janeiro Szövetségi Egyetemen. 2007–2012 között a Portugál Távközlési Intézettel (Portuguese Telecommunication Institute) működött együtt több, a kép- és videófeldolgozásra összpontosító kormányzati finanszírozású projektben, majd 2013-ban csatlakozott a Poznani Műszaki Egyetem multimédiás távközlési csoportjához, mint látogató posztdoktori kutató. 2010–2012 között a Leiriai Műszaki Egyetem adjunktusa volt Portugáliában. 2013-ban került az Ericssonhoz kutatómérnökként, ahol online és offline hardver és szoftverkódolási eljárások alapalgoritmusainak fejlesztésén dolgozott. 2018 óta a MediaKind vezető videótömörítési mérnöke, elsősorban az AI és ML alkalmazásával foglalkozik, médiafeldolgozó és -továbbítási eljárásokra összpontosítva. Ezek az eljárások a kódoló optimalizálástól az olyan új területekig terjednek, mint a szuperfelbontás és a videószemantikai elemzés.

BORDÁS CSABA 1993-ban szerzett mérnöki oklevelet a Kolozsvári Műszaki Egyetem gyengeáramú és távközlési szakán, majd 1993 és 2001 között Marosvásárhelyen dolgozott elsősorban az erdélyi analóg telefonhálózat digitalizációjában. Közben asszisztensként tevékenykedett az induló Gábor Dénes Főiskola helyi tagozatában. 2001-től csatlakozott a Ericsson csapatához Budapesten, először rendszertervezőként, majd a vezetékes ügyfélkör és a Magyar Telekom értékesítési ágazata műszaki igazgatójaként. Főleg FTTx-hozzáféréssel és IP-hálózati megoldásokkal foglalkozott, részt vett a hazai ADSL-, VDSL- és FTTH- hálózatok kiépítésében. 2015-től az Ericsson Deutsche Telekom globális értékesítési ágazat TV&Media igazgatója, ahol elsősorban a DT leányvállalatainak TV-megoldásaival, ezek fejlesztéseivel foglalkozott. 2018 óta hasonló szerepben a Deutsche Telekom ügyfélkör TV- és média-üzletágát koordinálja a MediaKind-nál.

HTE Diplomaterv és Szakdolgozat Pályázat, Almanach 2021

A HTE 2021-ben ismét meghirdette a már hagyományosnak mondható diplomaterv és szakdolgozat pályázatát, mely ebben az évben egy új kategóriával bővült. Így a mesterszakos (MSc) és az alapszakos (BSc) végzősök mellett idén először üzemmérnök alapszakosok (BProf) is pályázhattak. Ezzel párhuzamosan az IEEE Communication Society-vel és a Hungarian Joint ComSoc/MTT/AP/ED/EMC Chapter-rel közösen az angol nyelven megírt pályaművek számára „HTE – IEEE ComSoc Thesis Award" néven különdíj került kiírásra a korábbi évek hagyományait követve.

A pályaművek benyújtása és a bírálati folyamat teljes mértékben elektronikusan, transzparens módon, az Easy Chair elnevezésű konferencia-menedzsment rendszerben történt. A pályázat bírálatának lebonyolítását a HTE Tudományos Bizottsága felügyelte és ez a Bizottság tett javaslatot a díjazottakra a kialakult bírálati eredmények alapján.

A pályázatra idén rekordszámú, 40 pályamű érkezett, ebből 16 az MSc, 19 a BSc, és 5 a BProf kategóriában. A pályázók között öt felsőoktatási intézmény hallgatói képviseltették magukat; a BME mellett a Dunaújvárosi Egyetem, az egri Eszterházy Károly Egyetem, a Soproni Egyetem, valamint a győri Széchenyi István Egyetem hallgatói. Az IEEE különdíj versenyen kategóriától függetlenül az angol nyelven íródott pályaművek automatikusan részt vettek.

A pályaművek bírálatát egy felsőoktatási és ipari szakemberekből álló bíráló bizottság végezte. A bírálóknak értékelniük kellett a pályaműveket a témaválasztás korszerűsége, a kapcsolódó irodalom feldolgozása, saját munka mennyisége és színvonala, az elért eredmények, valamint a szerkesztés és formai elemek alapján. Ezen szempontokra adott bírálói értékelések összesítését követően állt elő a díjazottak sorrendje.

Ez évtől kezdődően szeretnénk minden évben a díjazott pályaművek egyoldalas kivonatait almanach formájában megjelentetni, hogy az utókor számára is bepillantást tudjunk nyújtani ezekbe az értékes, nívós munkákba. Ugyanakkor nem titkolt célunk, hogy ez a gyűjtemény érdekes és izgalmas korrajzként is szolgáljon, felvillantva az infokommunikációs területen a végzős hallgatókat megragadó, az adott időszakban korszerűnek számító témákat.

A jelen összeállítás a 2021-es *HTE Diplomaterv és Szakdolgozat Pályázat* díjazott pályaműveiből készült kivonatoknak és a szerzők rövid szakmai életútjának gyűjteménye.

Minden díjazottnak gratulálunk és sok sikert kívánunk további pályafutásukhoz!

Farkas Károly
HTE Tudományos területért felelős elnökségi tag, a pályázat koordinátora

További információk:

https://www.hte.hu/hte-diplomaterv-szakdolgozat-palyazat

TARTALOM

Mély megerősítéses tanulás szimuláció alapján valós környezetben önvezető járművekhez

ALMÁSI PÉTER BÉLA

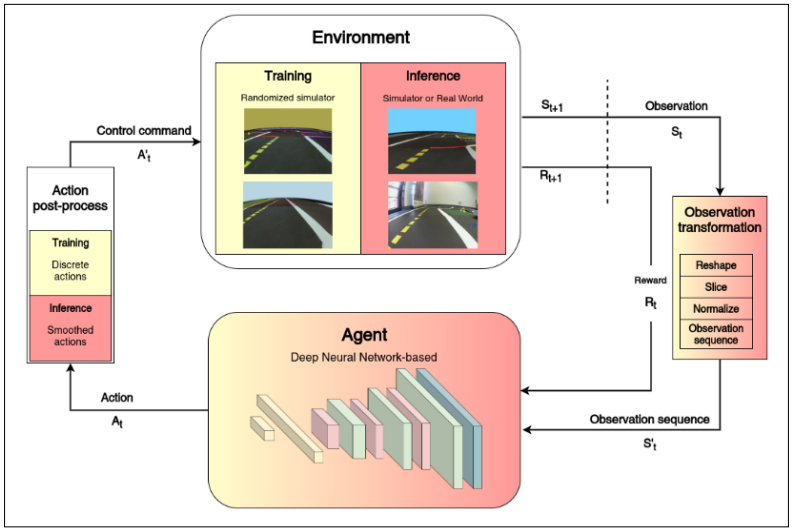
BME, Távközlési és Médiainformatikai Tanszék
almasipeti715@gmail.com

Konzulensek: Dr. Gyires-Tóth Bálint, Moni Róbert (BME, Távközlési és Médiainformatikai Tanszék)

Kulcsszavak: megerősítéses tanulás, mély tanulás, mesterséges intelligencia, neurális hálózatok, önvezető járművek

A mély neurális hálózatok kiemelkedő figyelemben részesültek az elmúlt években. Segítségükkel számtalan különféle alkalmazási területen sikerült minden korábbinál jobb eredményeket elérni, így például képfelismerési, objektumdetekciós, beszédfelismerési és -generálási, természetes nyelvfeldolgozási vagy idősorelemzési feladatokban is kiemelkedőnek bizonyultak. Ezek az eljárások sok esetben még az embernél is pontosabban oldják meg a számukra kijelölt feladatot.

A mély megerősítéses tanulás a gépi tanulási algoritmusok azon csoportját foglalja magába, amelyekben egy ágens neurális hálózatok használatával képes megtanulni egy környezetben egy bizonyos cél elérését a megfelelő akciók végrehajtásával. Ezzel a módszerrel vált lehetővé, hogy számítógépes algoritmusok legyőzzék a világbajnokokat különféle tábla- és számítógépes játékokban, például a Go-ban vagy a StarCraft II-ben.



Azonban a mély megerősítéses tanulás használata nagyobb kihívást jelent olyan feladatokban, amelyekben például valós robotokat vagy járműveket szeretnénk vezérelni. Ilyen feladatoknál jellemzően az ágenseket egy szimulátorban tanítják, majd a kész modellt „átültetik” a valós eszközre. A megerősítéses tanulás alkalmazása önvezető járművek esetében már szimulátorban is nehéz kihívás, hiszen ezek az algoritmusok jellemzően instabilak és nem rendelkeznek kellően megalapozott matematikai háttérrel. Továbbá a szimulátorban tanított ágensekre jellemző, hogy a valós környezetben történő használatkor jelentősen romlik a teljesítményük.

Dolgozatomban egy olyan mély megerősítéses tanuláson alapuló eljárást dolgoztam ki, amellyel lehetséges önvezető ágenseket tanítani szimulátor segítségével, és ezeket sikeresen át lehet ültetni valós járművekre is. Az ágensek a valós környezetben, valódi járműveken futtatva is hasonló pontosságot nyújtanak – a valós környezetből vett tanító minták nélkül.

A mellékelt *ábra* szemlélteti a kidolgozott módszer felépítését.

A dolgozatban két ágenst ismertettem, amelyeket különböző algoritmusok segítségével tanítottam. A módszerek ismertetése után összehasonlítottam és kiértékeltem az ágensek teljesítményét mindkét környezetben. Az eredményeket egy demonstrációs videóban is bemutattam, amelyre a hivatkozás megtalálható a dolgozatban.

A szerzőről

ALMÁSI PÉTER BÉLA tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetemen, mérnök-informatikus MSc szakon végezte. Diplomamunkájának fókusza a mesterséges intelligencia, ezen belül a mély megerősítéses tanulás, és az önvezető járművek területén végzett alkalmazásorientált kutatómunkát.

Protokoll állapotgépek visszafejtése

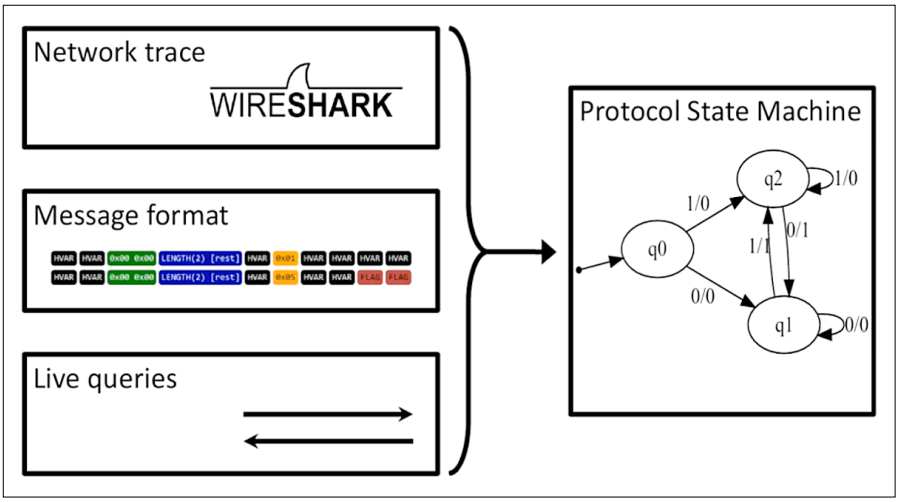
SZÉKELY GÁBOR

BME, Hálózati Rendszerek és Szolgáltatások Tanszék
szvgabor@gmail.com

Konzulens: Dr. Buttyán Levente (BME, Hálózati Rendszerek és Szolgáltatások Tanszék)

Kulcsszavak: automatikus protokoll-visszafejtés, formális nyelvtan inferálás, Mealy-automata

Az elmúlt 50 év során a számítógépek különböző technológiák segítségével egyre jobban összekötötté váltak, egyre növekvő hálózatokat alkotva. A számítógépek kommunikációját kommunikációs protokollok fejlesztésével tesszük lehetővé, melyek jelentős részének nyíltan nem érhető el a specifikációja. Ez hátráltatja a kompatibilis alkalmazások fejlesztését, a biztonsági tesztelést, illetve a hálózatot monitorozó és védő, a hálózatra csatolt hosztok védelmét is segítő szoftverek és házirendi szabályok létrehozását. Ismeretlen protokollokat használnak a hálózatuk koordinálására és parancsok kiosztására a botnetek is, amelyek olyan gépek hálózatai, melyek felett támadók vették át az irányítást, és melyeket többek között elosztott terheléses támadások indítására használhatnak. Ezen protokollokat a botnetek karbantartói hozzák létre és megértésük nagyban segítheti ezen hálózatok felszámolását.

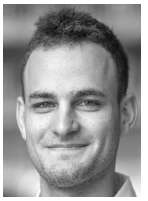


A fenti feladatok megoldásához nincs más választásunk, mint a protokollok visszafejtése. E feladat nehézségét a SAMBA-projekt látványosan demonstrálta, mely keretében az SMB-protokollt mintegy 12 év munkája során sikerült visszafejteni. Annak érdekében, hogy felgyorsítsák és megkönnyítsék az ilyen feladatokat, különböző automatikus protokoll-visszafejtési (APRE) módszereket fejlesztettek, megteremtve ezzel egy új kutatási területet. A legtöbb APRE-eszköz vagy a protokoll üzeneteinek szintaxisát, vagy a protokoll állapotgépét próbálja automatikusan megadni. Utóbbi azt írja le, hogyan követhetik a különböző típusú üzenetek egymást a kommunikáció során, különböző üzenetek fogadása esetén hogyan reagálnak a kommunikáló felek.

A dolgozatomban egy olyan algoritmust dolgoztam ki, mely képes automatikusan visszafejteni a kommunikációs protokollok állapotgépét. Az algoritmus magja az LM+ algoritmust használja fel, mely egy ismert algoritmus Mealy-automaták automatikus tanulására. Az LM+ felhasználása a hálózati üzenetek és az LM+ algoritmus által érthető be- és kimeneti betűk közötti oda- és visszatranszformálásával kerül felhasználásra. Emellett a lefedettség növelésének érdekében az algoritmus véletlenszerű, illetve vezérelt keresést végez olyan üzenetekért, amelyek még nem felfedezett működést produkálnak. Az eredményül kapott Mealy-automaták könnyebb értelmezhetősége érdekében egy utófeldolgozási algoritmust is definiáltam. Az algoritmusok implementációit tesztelésre egy kitalált protokollon és két valódi protokollon, a Modbuson és MQTT-n teszteltem.

Az implementáció a mellékelt *ábrán* szemléltetett módon működik: a futtatásához szükség van egy élő rendszerre, mely képes válaszolni a vizsgált protokollon küldött üzenetekre, illetve az üzenetek formátumára, hogy kategorizálni, valamint generálni lehessen helyes üzeneteket. Opcionálisan fel tud használni valós, rögzített hálózati forgalmat a futási idő rövidítésére és a nagyobb pontosság eléréséhez. Kimenete pedig olyan Mealy-automata, mely a vizsgált protokoll működését írja le.

A szerzőről



SZÉKELY GÁBOR tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetem mérnökinformatikus MSc szakán végezte. Tanulmányai során az IT-biztonság szakterületére fókuszált, melyen belül gyakorlati és kutatói munkát is végzett.

Hangforrások lokalizációja mikrofonrendszerekkel

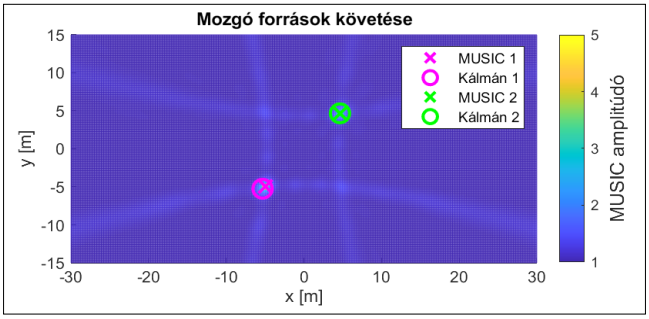
CSÓKA BENCE

BME, Hálózati Rendszerek és Szolgáltatások Tanszék
csokabence996@gmail.com

Konzulens: Dr. Fiala Péter (BME, Hálózati Rendszerek és Szolgáltatások Tanszék)

Kulcsszavak: akusztika, mikrofontömbök, nyalábformálás, Kálmán-szűrő, drónok

Munkám során a célom hangforrások pozíciójának meghatározása volt akusztikai kameraként használt mikrofontömbök segítségével. Ehhez két fő problémát kell megoldani, a fókuszálást és a forráslokalizációt. A fókuszálás során a tér egy adott irányából érkező hangot kiemeljük a Delay and Sum módszerrel, a mikrofonok vett jelének megfelelő erősítésével és késleltetésével. A forráslokalizáció során a hangforrás pozícióját becsüljük meg virtuális forráspozíciók közül (amik együttesen alkotják az akusztikus vásznat) a legvalószínűbbet választva. A két feladatot együttesen elvégezve végigpásztázzuk a tér egy részét, létrehozunk egy amplitúdótérképet és ez alapján végezzük el a becslést.



A vett jelek erősítését és késleltetését nyalábformáló algoritmusok (pl. CBF, MUSIC, CS) segítségével határozzuk meg. A CBF (Conventional Beamforming) egy nagyon egyszerű és gyors, de pontatlan módszer. A MUSIC (Multiple Signal Classification) a vett jelek keresztspektrum-mátrixának szétválasztásán alapul. A Compressive Sensing (CS) elven alapuló Compressive Beamforming a forráslokalizációt egy konvex optimalizálási problémára vezeti vissza. Egy ritkasági kényszerfeltételnek köszönhetően az amplitúdótérképen kevés nullától eltérő mező van, azaz a kamerakép nagy felbontású, de nagy a számításigénye.

Az akusztikus vászon pontjai általában egy síkon vagy egy gömbfelületen helyezkednek el. Egy ilyen vászon csak iránybecslésre alkalmas. A fókusztávolság változásával ugyanolyan mérési elrendezés esetén jelentősen változik a kamerakép minősége, így a vászon kiterjeszthető három dimenzióba a távolság meghatározása céljából. Egy elsődleges vásznon meghatározzuk az irányt, majd ebben az irányban egy egyenes mentén létrehozunk egy másodlagos vásznat, aminek a pontjai különböző távolságokra vannak a mikrofontömbtől. Ezzel a másodlagos vászonnal is végrehajtva a szükséges számítási lépéseket, már megbecsülhető a távolság.

A Kálmán-szűrő egy olyan algoritmus, ami optimális becslést ad változó rendszerek állapotára. Esetünkben ez a rendszer egy mozgó hangforrás, az állapota pedig a pozíciója és a sebessége. Az algoritmus a méréseken kívül a rendszer korábbi állapotát is figyelembe veszi, és ezek alapján végzi el a becslést, így a nyalábformáló algoritmusokat kiegészítve pontosabb végeredményt tud adni. Kiterjeszthető úgy, hogy nemlineáris rendszereket is kezeljen (pl. Unscented Kalman Filter), illetve, hogy egyszerre több hangforrást is kövessen.

Az algoritmusok MATLAB-környezetben lettek tesztelve szimulációkon és mérési eredmények feldolgozásán keresztül. A három nyalábformáló algoritmus két és három dimenzióban is hasonló eredményt adott, a CBF pontatlansága és a CS nagy számításigénye között a MUSIC-algoritmus jelentette az arany középutat. A Kálmán-szűrő a MUSIC-kal együttműködve képes mozgó források követésére (ahogy a mellékelt *ábrán* egy pillanatképen látható). Az algoritmus távolságbecslő része akkor és csak akkor képes helyes eredményt adni, ha az iránybecslés maga is helyes volt, és távoli források esetén (pl. 50 méterre a mikrofontömbtől) meglehetősen pontatlan.

Szabadtéri mérések során drónok voltak a hangforrások, amelyek pozícióját kellett becsülni. A körülmények itt kevésbé voltak kedvezőek, a zavarforrások kiszámíthatatlanabbak voltak a szimulációkhoz képest. Az iránybecslés a MUSIC-algoritmus és a Kálmán-szűrő segítségével megfelelően működött, néhány pillanatot kivéve, amikor valamilyen erős zavarforrás megnehezítette a becslést, illetve amikor a drón már túlságosan távol volt a mikrofontömbtől. A távolságbecslés által adott eredmény azonban nagyon instabil volt, így nem használható még mérési eredmények feldolgozására; ennek továbbfejlesztése fontos jövőbeni feladat.

A szerzőről

CSÓKA BENCE tanulmányai:
2007–2015: Gyöngyösi Berze Nagy János Gimnázium.
2015–2019: BME (BSc).
2019–2021: BME (MSc).

Csatornabecslés milliméteres hullámhosszú masszív MIMO rendszerekben

CSATHÓ BOTOND TAMÁS

BME, Szélessávú Hírközlés és Villamosságtan Tanszék
csatho.botond@edu.bme.hu

Konzulensek: Dr. Horváth Péter, Dr. Horváth Bálint Péter
(BME, Szélessávú Hírközlés és Villamosságtan Tanszék)

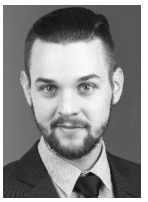
Kulcsszavak: vezeték nélküli kommunikáció, masszív MIMO, milliméteres hullámhossz, csatornabecslés, szoftverrádió

A masszív MIMO (massive Multiple-Input Multiple-Output) az egyik legkorszerűbb többantennás vezeték nélküli kommunikációs technológia, melyben a több tíz vagy több száz antennával rendelkező bázisállomás kommunikál a jellemzően egyetlen antennával ellátott felhasználókkal. A szóban forgó technológia megoldást jelent a korlátozott frekvencia-erőforrások melletti folyamatosan növekvő adatátviteli sebesség iránti igényre. A masszív MIMO sarokköve az ötödik generációs hálózatoknak, használatával a bázisállomás egy időben, azonos frekvenciasávban több mobil állomással is képes kommunikálni, azokat térben elválasztva. Igazolt, hogy a bázisállomás antennaszámát növelve nő a cellán belül elérhető eredő spektrális hatékonyság, így a jövőben akár a több száz vagy több ezer antennával szerelt bázisállomások fejlesztése várható.

A masszív MIMO működéséhez elengedhetetlen a rádiós csatorna ismerete, amit méréssel határoz meg a rendszer. Időosztásos működés esetén a rádiós csatorna reciprocitására támaszkodva a mérést elegendő felmenő ágban elvégezni, jelentősen csökkentve ez által a csatornabecslés időtartamát. Az adatsorozatok bázisállomás-oldali előkódolása és dekódolása mért csatornaparaméterek segítségével végezhető el, ebből fakadóan a csatornabecslés feladatköre kulcsfontosságú. A cellás rendszerekben használható csatornabecslési eljárások vizsgálata jelenleg is a tudományos érdeklődés központjában áll.

Dolgozatomban vizsgáltam a masszív MIMO-rendszerek csatornabecslését, ezt követően az ötödik generációs hálózat fizikai rétegéből kiindulva csatornabecslő keretrendszert terveztem. A kivitelezéshez szoftverrádiós platformot használtam, a rendszer teljesítőképességét szimulációval vizsgáltam és méréseket készítettem milliméteres hullámhossz-tartományban is, igazolva a jelfeldolgozási algoritmusok megvalósításának helyességét. A mellékelt *ábrán* egy validációs mérés látható, ahol 38,8 GHz-en, pont-pont összeköttetésben vizsgáltam a keretrendszer működését.

A szerzőről



CSATHÓ BOTOND TAMÁS tanulmányait a Debreceni Református Kollégium Dóczy Gimnáziumában, majd a Budapesti Műszaki és Gazdaságtudományi Egyetem (BME) Villamosmérnöki és Informatikai Karának villamosmérnök BSc- (2019, minősítés: kiváló) és MSc- (2021, minősítés: kitüntetéses) képzésein végezte. Jelenleg a BME Villamosmérnöki Tudományok Doktori Iskola doktorandusza, masszív többantennás kommunikációs rendszerek témakörében. Szakmai érdeklődési köre a vezeték nélküli kommunikáció, jelfeldolgozás és elektromágneses térszámítás tématerületei.

Elektromágneses hullámterjedés szimuláció plugin készítése QGIS szoftverhez

LUKÁCS BALÁZS

Széchenyi István Egyetem, Távközlési Tanszék
balazs.lukacs@rf.sze.hu

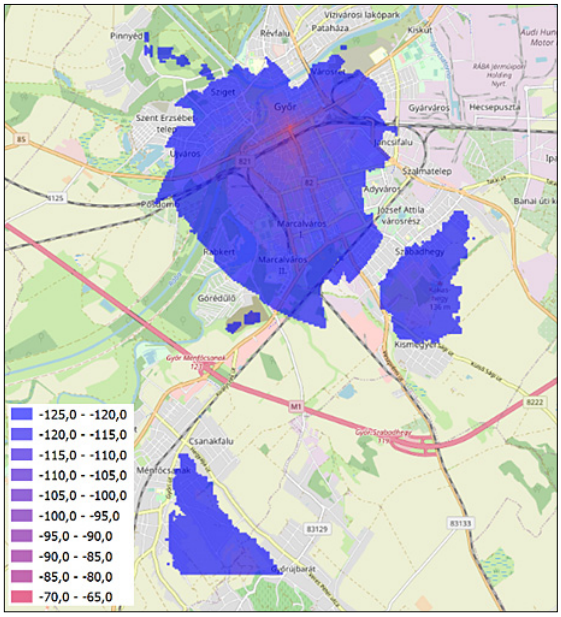
Konzulensek: Prukner Péter (Széchenyi István Egyetem, Távközlési Tanszék), Friedl Gergely (Jaguar Land Rover Hungary Kft.)

Kulcsszavak: 5G, hullámterjedés, GIS, szimuláció, 3,5 GHz

A dolgozatban bemutatásra kerül egy GIS-(Geographic Information System) szoftverhez készített plugin. A létrehozott plugin képes különböző empirikus hullámterjedési modelleket alkalmazni, hogy meghatározza a várható térerősség-értékeket a kijelölt területen belül.

A GIS-rendszereket kifejezetten helyhez kötött, aszszociált adatok tárolására, csoportosítására és feldolgozására fejlesztették ki. Ezáltal rendkívül sok területen alkalmazzák őket. Jelen esetben a hullám terjedésének vizsgálatához jelent egy nagyszerű eszközt, hogy a térbeli adatokat megfelelően lehessen kezelni és megjeleníteni.

A hullámterjedési modelleket már évtizedek óta használják a vezeték nélküli hálózatok előzetes lefedettségi vizsgálatához. Ez a módszer jelentős mennyiségű időt, erőforrást takarít meg a hálózattervezők számára. Folyamatosan fejlesztik a különböző hullámterjedési modelleket a minél pontosabb szimulációs eredmények eléréséhez, viszont a pontosabb eredményekhez egyre több és pontosabb adatra is szükség van. Ezért a diplomamunkában empirikus modellek kerültek bemutatásra és implementálásra a GIS- rendszerben. Az empirikus hullámterjedési modellek viszonylag kevés információval is pontos eredményeket szolgáltathatnak. Az ennél pontosabb lefedettség-szimulációkhoz már szükség van a terepviszonyokra vonatkozó adatokra (vegetáció, magasságértékek, talaj vezetőképessége) és egy vektoros épület-adatbázisra, az épületek anyagával együtt.



A diplomamunkában bemutatásra került a QGIS-szoftverhez készített plugin, valamint a plugin elkészítésének folyamata is, a különböző hullámterjedési modellek és a vezeték nélküli hálózatok tervezési folyamatai, a csillapítási tartalékok kiszámításának módszere. Végezetül pedig a létrehozott szimulációs szoftver által kiszámított lefedettség-térképek kerültek összehasonlításra annak függvényében, hogy milyen adatok és algoritmusok lettek felhasználva a lefedettség szimulációjához. A mellékelt *ábra* is egy ilyen szimulációs eredményt mutat Győr városában.

A diplomamunkában kiértékelésre és összehasonlításra kerültek a felhasznált elektromágneses hullámterjedési modellek, és egy példa-szimuláción keresztül bemutatásra került a QGIS-plugin működése, használata és végeredménye.

A szerzőről

LUKÁCS BALÁZS villamosmérnöki BSc-fokozatát 2018-ban, MSc-fokozatát pedig 2021-ben szerezte meg a győri Széchenyi István Egyetemen. Jelenleg a Széchenyi István Egyetem Multidiszciplináris Műszaki Doktori Iskola hallgatója. 2017 óta foglalkozik IoT-hálózatokkal és hullámterjedési modellek vizsgálatával. Jelenleg az egyetem Rádiófrekvenciás Vizsgáló Laboratóriumában dolgozik, ahol EMC- és RED-direktíva szerinti megfelelőségi vizsgálatokkal is foglalkozik, továbbá különböző kutatási és fejlesztési projektekben vesz részt.

Gépi tanulás alapú anomáliadetekció

VAJDA DÁNIEL LÁSZLÓ

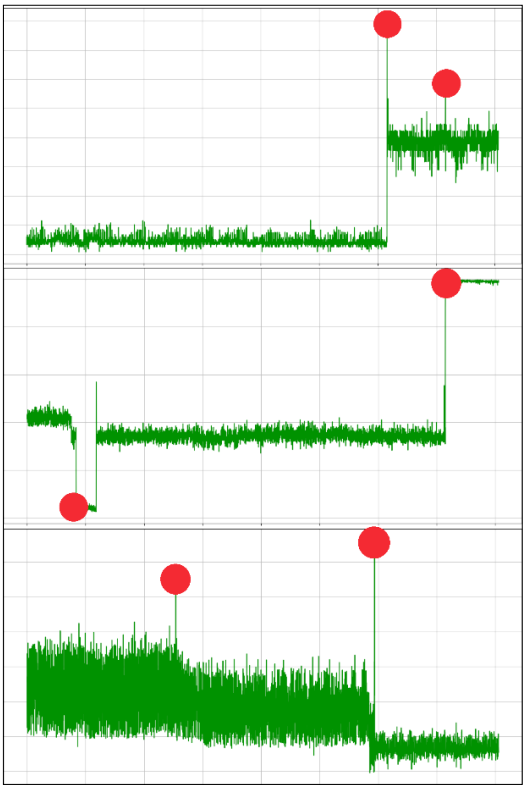
BME, Hálózati Rendszerek és Szolgáltatások Tanszék
vajdadaniellaszlo@edu.bme.hu

Konzulensek: Dr. Pekár Adrián, Dr. Farkas Károly
(BME, Hálózati Rendszerek és Szolgáltatások Tanszék / NETvisor Zrt.)

Kulcsszavak: anomáliadetekció, idősorok, LSTM, neurális háló, számítógépes hálózat

Ahogyan egyre több és több eszköz csatlakozik számítógépes hálózatokhoz, úgy lesz azok infrastruktúrája egyre komplexebb. Ezen hálózati eszközök, rendszerek és szolgáltatások folyamatos felügyelete manapság lényegesebb, mint valaha. Ennek számos haszna lehet, mint például üzemzavar előrejelzése; leállások elkerülése azáltal, hogy előre azonosítjuk azok jeleit; rendszerek teljesítményének monitorozása; továbbá rendszerek biztonságának felügyelete és az esetleges támadások észlelése.

Hagyományos módszerekkel azonban ezeket a funkciókat megbízhatóan, hatékonyan, valós időben megvalósítani korántsem egyszerű feladat. Ezt segíti elő a hálózati telemetria paradigmája, mely egy modern eljárás a hálózati eszközökből kinyerhető, idősor-alapú telemetriai adatok gyors, hatékony és automatikus begyűjtésére. A begyűjtött adatokat azonban fel kell dolgozni, hogy képesek legyünk detektálni a helytelen működésre utaló jeleket, amit gépi tanuló algoritmusok segítségével lehet hatékonyan megvalósítani. Ezt a folyamatot nevezzük anomáliadetekciónak.



A mellékelt *ábrán* három, hálózati eszközök processzor-teljesítményét megjelenítő idősor látható példaként, ezek forrása a Numenta Anomaly Benchmark platform. Piros pontok jelölik a valódi anomáliák helyeit az adatokban. Ezen pontok megtalálása jelenti a kihívást, a lehető legváltozatosabb forrású és jellegű idősorok esetén, eközben a lehető legkevesebb hamis jelzést produkálva.

Ez a dolgozat tehát kifejezetten az anomáliadetekcióra összpontosít, új megvilágításba helyezve annak idősor-alapú telemetriai adatokon történő használatát. Az irodalomkutatás során az ún. Long Short-Term Memory (LSTM) alapú, ReRe elnevezésű algoritmust azonosítottam, mint a jelenleg elérhető leghatékonyabb eljárást. Azonban a vizsgálataim azt mutatták, hogy még ez az eljárás is számos limitációval rendelkezik. Ezért a dolgozatban bemutatom az algoritmus általam továbbfejlesztett, Alter-Re^2-nek elnevezett változatát, melyben az eredeti eljárást az úgynevezett öregítés módszerével, illetve az adatok egy csúszóablakban való feldolgozásával egészítettem ki.

Egy, a bevezetett módszerek hiperparamétereit automatikusan beállító algoritmust is kidolgoztam. Az így elért teljesítményjavulás ígéretes, az Alter-Re^2 algoritmus átlagosan háromszor jobban, de legalább úgy teljesített, mint a ReRe, tíz különböző adatsoron végzett vizsgálat eredményei alapján.

Továbbá a dolgozatban kitérek arra, hogyan függ a ReRe és az Alter-Re^2 algoritmusok megbízhatósága és pontossága az elemzett adatsor típusától. Kategóriákba sorolom a feldolgozott adatsorokat az adatok mintázatai alapján, majd elemzem az algoritmus működését kategóriánként.

Meggyőződésem, hogy az Alter-Re^2 előnyösen használható számos területen, ahol gyors és pontos anomáliadetekcióra van szükség, mint például a hálózati telemetria, IoT-szenzorfolyamok, valamint behatolók, hibák és csalások észlelése esetén.

A szerzőről

VAJDA DÁNIEL LÁSZLÓ BSc-diplomáját a Budapesti Műszaki és Gazdaságtudományi Egyetemen szerezte meg 2021-ben, villamosmérnök szakon. Jelenleg ezen tanulmányait folytatja mesterszakos hallgatóként. Kutatásait többek között a NETvisor Zrt. gyakornokaként végzi.

Hálózati topológia bővítése regionális hibák védelmére

HAJDÚ ZSOMBOR LÁSZLÓ

BME, Távközlési és Médiainformatikai Tanszék
hajdu@tmit.bme.hu

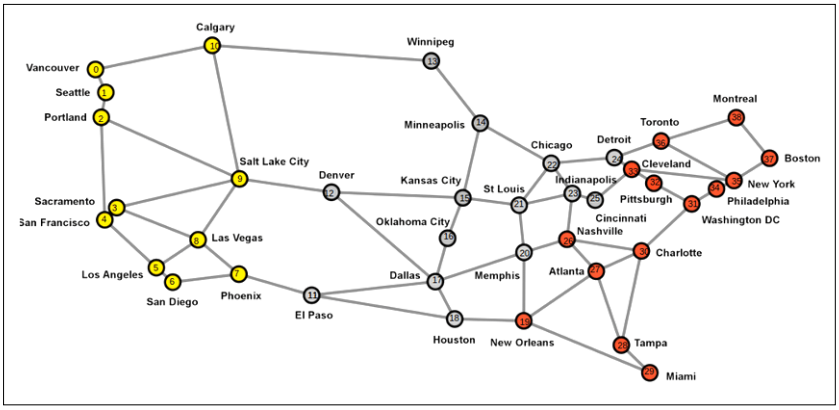
Konzulensek: Dr. Tapolcai János, Dr. Pašic Alija (BME, Távközlési és Médiainformatikai Tanszék)

Kulcsszavak: regionális hibák, hálózatvédelem, hálózattervezés, algoritmusok

Manapság az életünk egyre inkább internet-alapúvá válik, így a telekommunikációs hálózatok védelmének fontossága egyre relevánsabb lesz. A hálózatban történő kimaradások gyakran regionális szintű katasztrófák következményei, mint például egy földrengés, áradás, hurrikán vagy akár bombatámadás. A hálózatok kiterjedése miatt sokszor rengeteg olyan felhasználót is érint egy-egy hiba hatása, akik egyébként fizikailag messze vannak a katasztrófától.

A dolgozatomban a következő kérdésre kerestem a választ: Milyen módszerrel lehet olyan hálózatokat tervezni, amelyek ellenállóak regionális hibáknak, vagyis miként garantálható, hogy egy katasztrófa ne tudjon részekre szakítani egy hálózatot?

Esetünkben a problémát új összeköttetések kiépítésével oldjuk meg. Ekkor felmerül a kérdés, hogy az új optikai kábeleknek mely csomópontok közt, milyen útvonalon kell haladniuk, hogy optimális legyen a bővítés, persze amellett, hogy a hálózat „bombabiztosságát” – azaz azt, hogy a közvetlenül nem érintett felhasználók ne maradjanak szolgáltatás nélkül – is garantálni lehessen.



A szakdolgozatomban kitérek a problémát eddig (részben) tárgyaló szakirodalomra és az alkalmazott módszerekre, majd felépítem a matematikai modellt, amelynek segítségével a probléma formalizálható lesz: a hálózatot egy irányítatlan geometrikus gráfként *ábrázoljuk*, majd számítógépes geometriai módszerekkel meghatározunk bizonyos „veszélyzónákat”, amiket, ha katasztrófa (például egy bizonyos zónába eső epicentrumú földrengés) érne, a hálózat több részre esne szét. Ezután a feladat részekre bontásával és geometriai algoritmusok segítségével keressük az így keletkező komponensek közötti új élek legolcsóbb útvonalait.

Több lehetséges algoritmust dolgoztam ki a leginkább költséghatékony hálózatbővítés kiválasztására. Az új élekkel kiegészített, bővített hálózat már adott méretű regionális katasztrófáktól védett lesz, akárhol is történjenek azok. A különböző módszereket implementáltam, majd a hatékonyságukat összevetve elemeztem őket. A gördülékeny munkához és teszteléshez egy elosztottan működni képes rendszert is felépítettem.

A szerzőről

HAJDÚ ZSOMBOR LÁSZLÓ tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karán folytatta, ahol 2021 januárjában szerezte meg mérnökinformatikus végzettségét. 2018 óta a HSNLab kutatási segédje.

Objektumok felismerése és követése felhőből vezérelt drónokkal

CZURKÓ DÁNIEL

BME, Távközlési és Médiainformatikai Tanszék
czurko.dani@edu.bme.hu

Konzulens: Dr. Fehér Gábor (BME, Távközlési és Médiainformatikai Tanszék)

Kulcsszavak: Docker, felhőalapú robotika, objektumfelismerés, ROS, szimuláció

A drónok alkalmazásánál egy gyakran előkerülő megoldandó probléma, hogy a drónoknak követniük kell valamit vagy valakit. Így légi felvételek esetében gyakran egy mozgó embert, például egy biciklist vagy gördeszkást, vízi sportoknál szörföst vagy vitorlást, télen pedig egy síelőt. A csomagszállításnál is számos olyan szituáció adódhat, amikor valamilyen objektumot kell követnie a drónnak, mint például egy előre meghatározott objektumra való leszállásnál.

Robotok és drónok építése során az egyik fontos kérdés, hogy mekkora számítási kapacitást tervezzünk a robotunkra. Természetesen azt szeretnénk, ha minél több mindenre képes lenne, ehhez azonban nagy számítási kapacitást kellene a robotnak magával cipelnie, ami sokszor nem megoldható. Ugyanakkor, ha rendelkezésünkre áll egy nagyon kis késleltetésű hálózat – például egy 5G-s mobilhálózat –, és a hálózat szélén egy nagy számítási kapacitású számítógép – tipikusan valamilyen felhő-infrastruktúra formájában –, akkor meg lehet oldani, hogy szinte végtelen számítási kapacitás álljon a robotunk rendelkezésére. Ilyenkor az irányításához szükséges komplex és nagy számítási kapacitású vezérlési logikát kiszervezzük a hálózat szélén elhelyezkedő nagy számítási kapacitású számítógépbe.

A szakdolgozat során a feladatom egy felhőben futó objektumkövető rendszer elkészítése volt, mely képes egy szimulált drón kamerakép alapján történő irányítására oly módon, hogy a drón kövessen egy előre meghatározott mozgó objektumot. Erről látható egy pillanatfelvétel a mellékelt *ábrán*. A kép a drón kamerájának feldolgozott képét mutatja a felismert emberrel, valamint a függőleges és vízszintes segédvonalakkal. A drónnak a célja, hogy mozgásával a felismert embert a függőleges és vízszintes vonalak találkozásánál kialakult téglalapba irányítsa. Az ábra bal felső sarkában látható szöveg pedig az objektum-felismerés sebességét mutatja meg.

A feladat megoldása során mozgó objektumnak egy járkáló embert választottam. A rendszer tervezése és megvalósítása oly módon történt, hogy a szimulált drónt egyszerűen ki lehet cserélni egy valóságos drónra, valamint nem csak mozgó ember felismerésére és követésére alkalmas, hanem több mint ötven másik mozgó objektumot is képes követni.

A szerzőről

CZURKÓ DÁNIEL a Szent István Gimnáziumban szerezte érettségi bizonyítványát, majd tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karának IMSC programjában folytatta. Jelenleg az egyetem MSc-hallgatója. Az egyetemen főként hálózatokkal és virtualizációval foglalkozik, szeret robotokat építeni és programozni.

Adatalapú döntéshozatal gépi tanulás eszközökkel tőzsdei adatokon

CSANAKI RICHÁRD

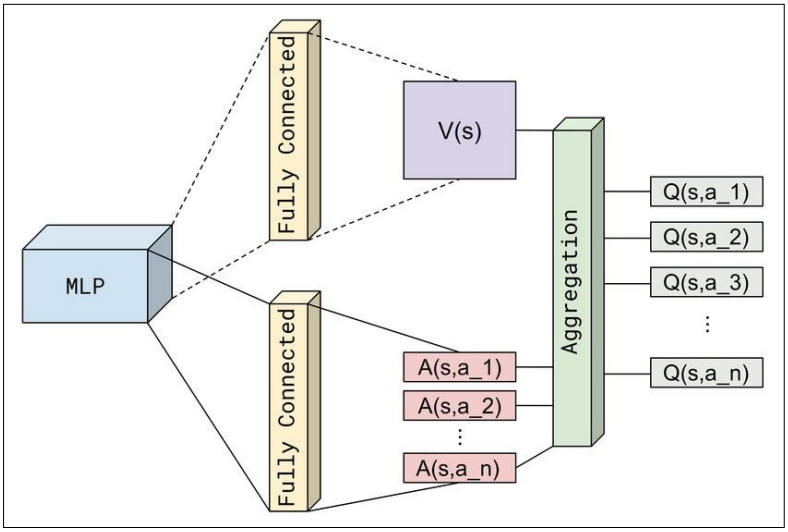
Soproni Egyetem, Informatikai és Gazdasági Intézet
csanakir@gain.uni-sopron.hu

Konzulens: Dr. Pödör Zoltán (Soproni Egyetem, Informatikai és Gazdasági Intézet)

Kulcsszavak: mesterséges intelligencia, visszacsatolásos mélytanulás, Q-learning, tőzsde

A szakdolgozatban bemutatom a gépi tanulás rövid elméleti hátterét és részletesen foglalkozom a tudományág egy részterületével, a visszacsatolásos tanulással. Ennek során a használt ágens feladata, hogy egy Markov-láncként formalizált környezetben hozzon döntéseket, miközben egy előre definiált, doménspecifikus jutalom-függvényt hosszú távon maximalizál.

A jutalomfüggvény kimeneteivel lehet értékelni az úgynevezett Q-függvényt, amely a visszacsatolásos tanulásban használatos policy (π) függvény egy fajtája. Paraméterei a vizsgált környezet aktuális állapota, valamint az aktuálisan végrehajtandó lépés, így definiálva az ágens számára, hogy adott szituációban, állapotban milyen lépést tegyen.



A Q-függvényeket használó Q-learning módszer alapvetően egy modell nélküli tanulási megközelítés, azonban a dolgozatban ennek egy továbbfejlesztett verziójával dolgoztam, amelyben kettő neurális hálózat párhuzamosan tanul egymástól, a kettő modell közötti kommunikációt biztosító ágens segítségével.

$$Q^A_{t+1}(s_t,a_t) = Q^A_t(s_t,a_t) + \alpha_t(s_t,a_t)\left(r_t + \gamma Q^B_t\left(s_{t+1}, \arg\max_a Q^A_t(s_{t+1},a)\right) - Q^A_t(s_t,a_t)\right)$$

$$Q^B_{t+1}(s_t,a_t) = Q^B_t(s_t,a_t) + \alpha_t(s_t,a_t)\left(r_t + \gamma Q^A_t\left(s_{t+1}, \arg\max_a Q^B_t(s_{t+1},a)\right) - Q^B_t(s_t,a_t)\right).$$

A módszer egy gyakorlati feladatban került implementálásra, ahol két hat rétegű „Long Short-Term Memory" architektúrájú neurális hálózat tanul egymástól azért, hogy meghatározzanak egy optimális értékpapír-befektetési stratégiát a bemeneti, két attribútumból (zárási árfolyam, napi kereskedett volumen) álló adathalmaz alapján.

A gyakorlati feladatban több megközelítés is kipróbálásra került, így az egyéni részvényekkel való kereskedés, „index fund"-részvényekkel való kereskedés, valamint azonos szektorba tartozó részvényekkel történő kereskedés azért, hogy egy modellt lehessen alkalmazni több részvénnyel lefolytatott ügylethez is. Az egyéni részvények esetén stabil profitráta érhető el, míg azon adathalmazokon, amelyek több részvényt tartalmaztak, volatilis hozamok mutatkoztak.

A szerzőről

CSANAKI RICHÁRD tanulmányait a Soproni Egyetem Simonyi Károly Műszaki, Faanyagtudományi és Művészeti Kar gazdaságinformatikus BSc szakán végezte, ahol 2021-ben kiváló minősítésű oklevelet szerzett. Tanulmányi, gazdaságinformatikus szakmai, Új Nemzeti Kiválóság Program és Nemzeti Felsőoktatási ösztöndíjas hallgató, az Országos Gazdaságinformatikus Konferencia kitüntetett előadója, az Informatikai és Gazdasági Intézet volt kutatási asszisztense. Az egyetemi tanulmányok mellett a budapesti IT startup világban tevékenykedett, mint az Abylon Consulting Kft. gyakornoka, jelenleg a Magna International üzleti intelligencia fejlesztője Grazban, a Microsoft által tanúsított adatelemző. Kutatási területe a visszacsatolásos mélytanulás alapú mesterséges intelligencia rendszerek.

A „cloud native” alkalmazásfejlesztés hatékony praktikái

KERTÉSZ DÁVID RICHÁRD

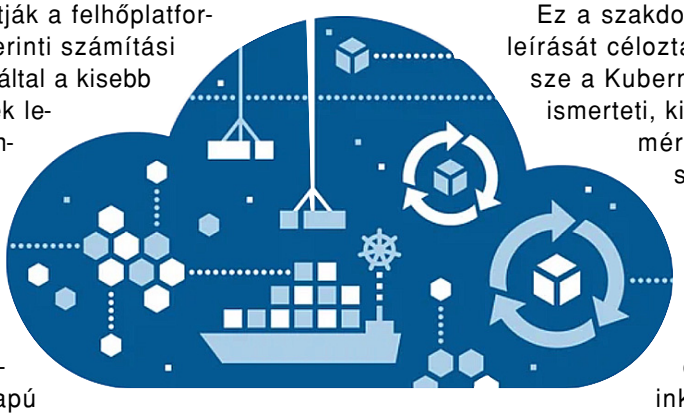
BME, Hálózati Rendszerek és Szolgáltatások Tanszék
david.r.kertesz@outlook.com

Konzulensek: Szabó Gergely (Origoss Megoldások Kft.)
Dr. Farkas Károly (NETvisor Zrt. / BME, Hálózati Rendszerek és Szolgáltatások Tanszék)

Kulcsszavak: Cloud Native, DevOps, felhőalapú fejlesztés, Kubernetes, rendszer monitorozás

Napjainkat a technológiai fejlődés exponenciális gyorsulása jellemezi. Nem ritka, hogy amire egy nagyobb sztenderd megveti a lábát a szakmában, máris versenyeznie kell akár több, gyorsabb működést, nagyobb költséghatékonyságot, vagy egyszerűbb használhatóságot ígérő, fejlődő technológiával.

A felhőalapú számítástechnika fejlődése felgyorsította a felhőalapú megoldások integrálását a szoftverfejlesztési és üzemeltetési folyamatokba. Azon vállalatok, amelyeknek nincsenek meg a megfelelő erőforrásaik arra, hogy a világ bármely pontján biztosítsák szolgáltatásaikat, felhasználhatják a felhőplatformok által nyújtott igény szerinti számítási erőforrások lehetőségét. Ezáltal a kisebb vállalatok is versenyképesek lehetnek a nagyobbakkal szemben és nyereségesek is maradhatnak az üzemeltetési költségeik óvatos menedzselésével.



Ez a folyamat – amiben a vállalatok elmozdulnak a saját maguk által fenntartott szerverektől a felhőalapú rendszerek irányába – helyet adott több olyan technológiának és metodikának, amelyek jelenleg a szoftverfejlesztést és üzemeltetést alapjaikban meghatározzák. A DevOps egy ilyenfajta kulturális folyamat, magába foglalva több agilis gyakorlatot, metodikát és automatizmust, amelyek a modern szoftverfejlesztés részei. Ezek összességükben napi szintről órákra csökkentik a szoftver-beüzemelés idejét, míg a minőségellenőrzés szempontjai ugyanúgy teljesülnek.

Az ezen gyakorlatok által meghatározott fejlesztési folyamatok is nagy mértékben automatizáltak. A folytonos folyamatok (continuous practices) meghatározzák az IT-csapatok napi munkavégzését. A csapatok gyors reagálási idejét részben a konténerizációs technológiák használata alapozza meg, melyek segítségével egy absztrakciós réteg kerül a szoftver és az azt futtató környezet közé.

Ezáltal a több konténerből álló rendszerek orkesztrációja fölé is lehet vonni egy absztrakciós réteget, amire az egyik bevett technológia a Kubernetes. A Kubernetest nagy mértékben kezdték el használni, miután a Google nyílttá tette a forráskódját, és manapság már vezető technológia a felhőalapú üzemeltetésben.

Az előbbiekben említett folyamatok a szoftverfejlesztésben és üzemeltetésben, valamint a fejlett felhőalapú technológiák a részeit képezik annak, amit úgy hívunk, hogy „cloud native”. A „cloud native” ökoszisztéma több, egymással kapcsolatban álló technológia szorosan összefüggő halmaza, melyek célja, hogy nyílt forráskódú megoldásokat biztosítsanak a fejlesztési és üzemeltetési folyamatok tisztán felhőalapú rendszerbe való átültetésére.

Ez a szakdolgozat ezen ökoszisztéma leírását célozta meg. A dolgozat első része a Kubernetes alapvető működését ismerteti, kitérve arra, hogy ez milyen mértékben változtatta meg a számítástechnika világát. A második rész egy gyűjtemény a fontosabb fejlesztési és üzemeltetési metodikákból és gyakorlatokból, amelyeket a modern IT-csapatok az egyre inkább online világban való hatékonyság és produktivitás okán követnek.

A harmadik rész pedig bemutatja azt a Kubernetes alapokon működő, komplex „cloud native” rendszert, amelyben egy egyszerű, a leírt gyakorlatokat követő folyamattal fejlesztett alkalmazás fut, melynek a célja, hogy tesztelhető legyen a működésén keresztül a felhőrendszer hatékonysága. Egy teszteset is leírásra kerül, mely példaként szolgál a teljes „cloud native” rendszer működésére éles környezetben.

A szerzőről

KERTÉSZ DÁVID RICHÁRD tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karán végezte üzemmérnök-informatikus szakon. Jelenleg az IBM-nél dolgozik szoftverfejlesztőként.

Energiaszolgáltatások felhasználóinak fogyasztásvizsgálata gépi tanulási módszerekkel

BABÁK BOTOND

BME, Távközlési és Médiainformatikai Tanszék
botond.babak@gmail.com

Konzulensek: Hegedüs Ákos (Sagemcom Magyarország Kft.)
Dr. Toka László (BME, Távközlési és Médiainformatikai Tanszék)

Kulcsszavak: gépi tanulás, csalásfelderítés, energiaszolgáltató, fogyasztásvizsgálat

A víz-, gáz- és áramszolgáltatók a technológia fejlődésének, az úgynevezett okos mérők megjelenésének köszönhetően egyre hatékonyabban tudják azonosítani azon felhasználóikat, akik szolgáltatásukat jogtalanul, vagy nem a mért mennyiségben használják fel. Ezen csalók sikeres felismerésében a sűrű mintavételezésen és a pontos helymeghatározáson túl a gépi tanulási módszerek is segítségünkre lehetnek.

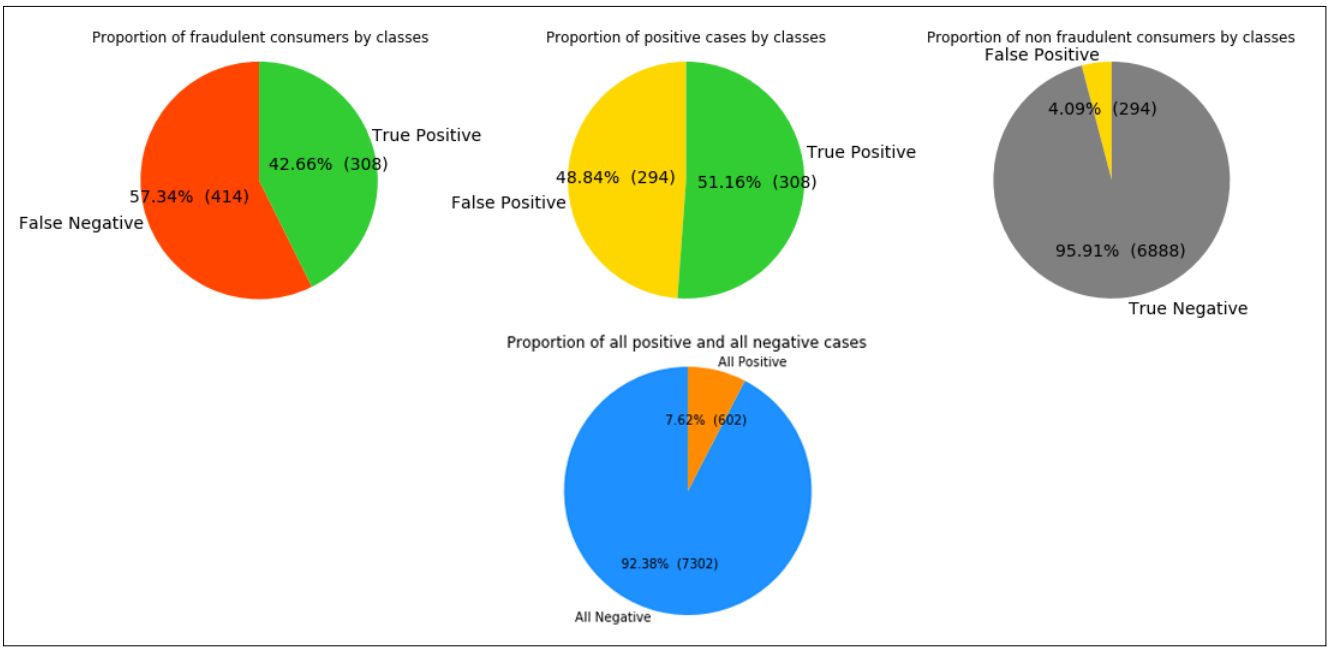
A dolgozatom során ennek a problémának a megoldására kerestem minél alkalmasabb gépi tanulási módszereket, illetve összehasonlítottam és úgy hangoltam azokat, hogy minél kedvezőbb eredményeket érjek el. Az összehasonlítás szempontjai közé tartozott, hogy az algoritmus esetlegesen a későbbiekben használható legyen ipari környezetben is. Emiatt az algoritmus betaníthatóságát, erőforrásigényét és költséghatékonyságát is figyelembe kellett venni.

A publikusan elérhető adathalmazok felhasználásával történt vizsgálataimban szereplő algoritmusok közül a legjobban teljesítő eljárás által elért eredmények láthatók mellékelt *diagramokon*. Így a csalók 42,66%-át megtalálta az algoritmus (első diagram), amihez a teljes fogyasztóbázisnak csak 7,62%-át kellett megvizsgálni (alsó diagram). Emellett a valóban megvizsgált fogyasztók közül is minden második ténylegesen csaló is volt (középső diagram), ami azt jelenti, hogy további adathalmazok építéséhez is alkalmazható az algoritmus, ezzel javítva a későbbi eredményeken.

Ennél sokkal radikálisabb megoldások is születtek, amik üzletileg nem feltétlenül érték volna meg, de a későbbiekben egy jobb adathalmaz felépítése után azokban is lehet potenciál.

A szerzőről

BABÁK BOTOND tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karán, üzemmérnök-informatikus szakon végezte. Szakmai tapasztalata:
2020–2021: Sagemcom Magyarország Kft. – Szoftvertesztelő/Adatelemző gyakornok.
2021–től: Accenture Magyarország Kft. – Azure Data Engineer.



Szeparált eszközmenedzsment hálózat fejlesztése a Richter Gedeon informatikai hálózatán

OLÁH MÁRK

BME, Hálózati Rendszerek és Szolgáltatások Tanszék
mark.olah.om@gmail.com

Konzulensek: Boór András (Richter Gedeon Nyrt.)
Dr. Holczer Tamás (BME, Hálózati Rendszerek és Szolgáltatások Tanszék)

Kulcsszavak: hálózatmenedzsment, „Out Of Band”, útválasztás, hálózattervezés

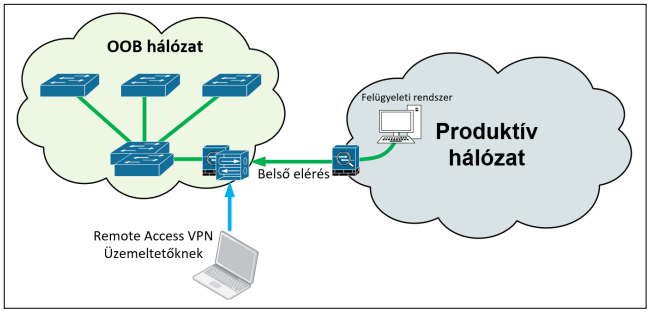
A modern informatikai rendszerek komplexitása folyamatosan nő. Ez azt is jelenti, hogy az azokat üzemeltető és fejlesztő szakembereknek egyre nagyobb kihívást jelent, hogy emellett a folyamatos bővülés mellett megőrizzék azt a kontrollt, olyan mértékű áttekinthetőséget, ami a kisebb rendszerek esetén természetesen jelen van.

A kontroll megtartásának egyik talán legszemléletesebb példája a hálózatban található eszközök menedzselhetőségének kérdése. Ezt támasztja alá, hogy a nagyméretű, akár több ezer eszközt magába foglaló hálózatok felügyeletével, menedzselhetőségével és annak megbízható működésével szemben épp olyan, vagy talán még magasabb elvárásaink vannak, mint a kisebb hálózatokkal szemben.

A nagyobb vállalatok – mint amilyen a Richter Gedeon Nyrt. is – sok esetben több telephellyel rendelkeznek, azokon belül pedig több különálló szerverhelyiséggel. Ez az elosztott struktúra megbízhatóbbá teszi az üzemeltetést, azonban kihívást jelent a hálózati eszközök kezelésének szempontjából. A legkönnyebben kialakítható és a leginkább költséghatékony, ugyanakkor a legkevésbé robosztus megoldás, amikor a hálózat már használatban lévő kapcsolatait használjuk erre a célra, és az üzleti forgalom csatornáin keresztül valósítjuk meg a menedzsment funkcióit. A vállalat továbbfejlesztette ezt a kialakítási sémát oly módon, hogy a kulcsfontosságú eszközei mellé külön hálózati kapcsolókat telepített a menedzsment-forgalom továbbítására.

A szakdolgozat-feladatom a meglévő menedzsmenthálózat továbbfejlesztése volt egy olyan megoldássá, ahol az eszközök elérésével és megfigyelésével kapcsolatos forgalom teljes mértékben elkülönül az üzleti forgalomtól. Ennek létrehozásához felmértem és meghatároztam a már meglévő menedzsmenthálózat azon központi eszközeit, amelyeket felhasználva a lehető legkevesebb topológiaváltozással elérhetem a kívánt eredményt. A szükséges információk birtokában készítettem el azt a hálózati tervrajzot, amely a vállalat teljes eszközmenedzsment hálózatát szemlélteti, és amely az új OOB- (Out Of Band) menedzsmenthálózat terveinek alapjául szolgált. A tervezési fázis további szakaszaiban azonban egyetlen átfogó tervrajz helyett inkább több különböző szintű leképezését készítettem el a kialakítandó konstrukciónak. Az egyes nézetek az OSI-modell szerinti rétegeknek felelnek meg, kezdve a fizikai rétegtől egészen a harmadik, azaz a hálózati rétegig. Ez a tervezési módszer nagyon hasznosnak bizonyult, mivel így könnyedén átláthattam az egyes rétegekkel kapcsolatos követelményeket és kihívásokat. A fizikai réteget leíró tervrajz segített az összeköttetések és az azokhoz szükséges szerelési anyagok felmérésében, míg a második réteg nézete alapján könnyedén meghatározhattam a VLAN-ok hatóköreit. Végül pedig a harmadik réteget leíró tervrajz segített az új OSPF-area megtervezésében. Így jobban átláthattam, mit jelent az útválasztásra vonatkozóan leválasztani egy kisebb menedzsmentszigetet a produktív hálózatról, és hozzácsatolni az új OOB-menedzsmenthálózathoz. A tervezési eljárás nyomán eszközölt változtatások eredménye az *ábrán* látható.

Az új topológia megbízható menedzsment célú elérést biztosít a produktív hálózaton fellépő hibák esetén is. Másik előnye, hogy a hálózathoz való hozzáférés könnyebben felügyelhető és szabályozható.



Záró gondolatként szeretném megjegyezni, hogy ez a projekt számomra sokkal többet jelentett, mint a szakdolgozatom témáját adó feladat. Amikor ezen dolgoztam, rengeteg tapasztalatot szereztem nem csak a tervezés, de az eszközök konfigurálásának, a feladatok ütemezésének és a lehetőségek számbavételének kapcsán is.

A szerzőről

OLÁH MÁRK tanulmányait a BME Villamosmérnöki és Informatikai Karán végezte. Szakmai gyakorlatát a Richter Gedeonnál töltötte a Network and Security csoport tagjaként, ahol routing and switching, VoIP, tűzfalak és DWDM-kapcsolatok témakörökkel foglalkozott. Oklevelének megszerzése után a Unisys Magyarország Kft.-nél Network Design Engineer pozícióban helyezkedett el, ahol leginkább magasszintű hálózati tervek készítésével és hálózati projektek támogatásával foglalkozik.

TDoA-Based Indoor Positioning Over Cellular 5G Network

TDoA alapú beltéri helymeghatározás cellás 5G mobilhálózaton

PAPP ZSÓFIA

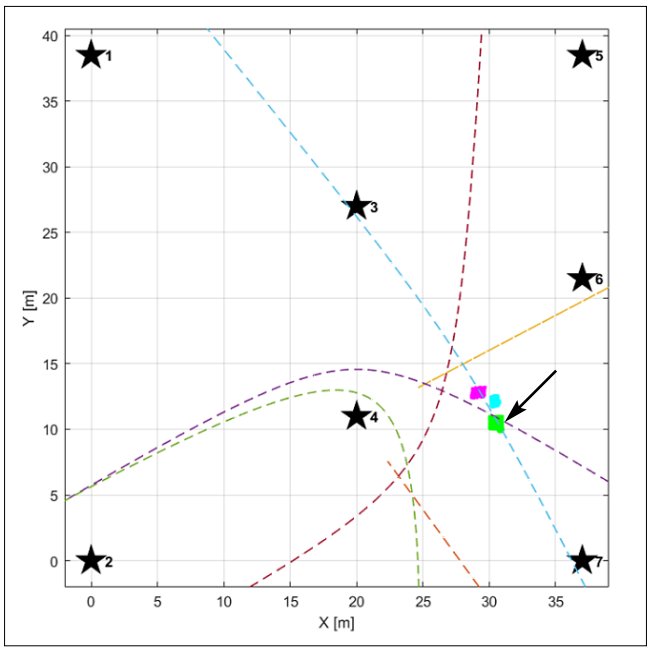
BME, Távközlési és Médiainformatikai Tanszék
zsofi.bme@gmail.com

Konzulensek: Dr. Pašic Alija (BME, Távközlési és Médiainformatikai Tanszék)
András Dániel (Ericsson Magyarország Kft.)

Kulcsszavak: 5G mobilhálózat, beltéri helymeghatározás, beltéri jelterjedési modell, csatornamodell, NLOS kiküszöbölés

Az 5G-hálózatok fejlődése és elterjedése várhatóan jelentős átalakulást fog eredményezni életünk számos területén. Az iparban, a közlekedésben, az egészségügyben, az energiagazdálkodásban is lényeges fejlődésre számíthatunk a mobilhálózatok új generációjának köszönhetően. Az új alkalmazási területeken számos esetben elengedhetetlen a mobilhálózathoz kapcsolódó eszközök pontos és megbízható helymeghatározása, gondoljunk csak az önvezető járművekre vagy az intelligens robotokra. Nyílt területeken jól bevált megoldást kínál a GPS, beltéri pozicionáláshoz azonban új módszerekre van szükség.

Egy ilyen megoldás kidolgozása során a legnagyobb kihívást a beltéri jelterjedési sajátosságok jelentik: a rádiójelek terjedését gyakran nehezítik falak és egyéb akadályok, melyek a jelek csillapítását, visszaverődését és szóródását okozzák. Ezen hátrányos hatások miatt nehéz olyan beltéri helymeghatározó rendszert alkotni, amely elfogadható telepítési és üzemeltetési költségek mellett képes az elvárt pontosságot is biztosítani. Egy ígéretes megoldási alternatíva az 5G-mobilhálózaton keresztül történő pozicionálás, mivel megbízható, eszközfüggetlen helymeghatározást tesz lehetővé járulékos telepítési költségek nélkül.



Diplomamunkámban egy beltéri, 5G-alapú mobilpozicionálást megvalósító rendszer tervezésének első lépésein haladtam végig. Az elméleti háttér részletes felderítését követően egy szimulátorprogramot készítettem a jövőbeli rendszer várható teljesítményének előrejelzésére. A szimulátorban valószínűségi változókkal modelleztem a különböző hibaforrásokat, melyek közül a legérdekesebb az NLOS- (Non-Line of Sight) terjedésből adódó időmérési hiba. Ennek szimulációjához többféle beltéri jelterjedési modellt implementáltam. A szimulált mérési értékeket felhasználtam a helymeghatározást végző algoritmusok optimalizálásához, melynek során sikerült a helymeghatározás hibájának átlagát 3,6 méterről 2,7 méterre csökkentenem.

A mellékelt *ábra* a szimulátor egyik kimenetét mutatja, amely két különböző algoritmus által becsült pozíciókat (tónusos pontok) hasonlít össze adott rendszerjellemzők mellett. Fekete csillagok jelölik az 5G referenciaantennák helyzetét, a hozzájuk tartozó TDoA (Time Difference of Arrival) méréseket hiperbolák szemléltetik. A szimuláció során a mobil tényleges pozíciója egy 1m*1m nagyságú (nyíllal jelölt) négyzeten belül helyezkedik el.

A szerzőről

PAPP ZSÓFIA tanulmányait a Budapesti Műszaki és Gazdaságtudományi Egyetemen végezte, ahol villamosmérnök alapképzésben vett részt és a mobilhálózatok irányába specializálódott. Szakmai tudását a mesterképzés alatt tovább mélyítette, miközben az Ericsson Network Location-nél végzett munkája során gyakorlati tapasztalatot is szerzett az 5G-hálózatok fejlesztésében. Főbb érdeklődési területei a pozicionálási módszerek és algoritmusok, illetve a beltéri jelterjedési szimulációk és modellek.

Summaries • of the papers published in this special issue

This Special Issue is compiled from the papers of the 23nd HTE Infokom 2021, the Infocommunications Networks and Application Conference, organized by the Scientific Association for Infocommunications (HTE).

Recent developments in service provider architectures in the age of cloud computing and cloud based applications

Keywords: SDN, merchant silicon, IPoDWDM, cloud computing, container based virtualization

Service providers has to deal with ever increasing bandwidth consumption (driven by Wi-Fi, 5G, and other broadband technologies) and the adoption of cloud-based applications. Not only the amount of bandwidth, but also the traffic patterns are changing. On the other side, network development must happen despite shrinking CAPEX and OPEX conditions. Simplification and automation is needed in the network. The paper explains the most important aspects of the evolution of the service provider networks.

The emergency service during the COVID-19 pandemic

Keywords: emergency communication, users needs, public utility, answers to the challenge

The paper presents the over 15 years of governmental purpose telecommunication service provider's difficulty during the pandemic in EDR public utility maintenance with the users needs. The COVID-19 pandemic caused increased extra tasks within our users which means that the EDR provider had to answer these needs due to short amount of time to maintain the EDR public service operating smoothly. Above all, the paper presents the institutional, regime and administrative measures in favor of the business continuity of Pro-M Zrt.

Edge Computing as a disruptive technology

Keywords: Edge Computing, 5G, IoT, industry 4.0, mobile network services

Edge Computing is selected as one of the top popular technologies by many market and technology surveys. Although the concept has been there for a decade or more, 5G and IoT system have brought it to the spotlight. This paper introduces some aspects of its disruptive nature. After a short introduction of the edge computing use cases, we will discuss the roles and intentions of different players in the possible edge computing business value chains. Then we discuss technical solutions and highlight some key technical problems yet to be solved.

What does the 5G network offer to the average consumer?

Keywords: 5G networks, 5G NSA and 5G SA system, average consumer, new 5G services

The paper looks for the answer to what 5G networks – while promising many benefits for vertical industries – offer to average consumers now and in the future. The first part of the paper presents the current market situation of 5G systems and then discusses the difficulties of market introduction and the solution of the „chicken or the egg" dilemma that has arisen here. The next section introduces the applications that can already be used under the 5G non standalone (5G NSA) system, and then looks to the future with the versatile features of the 5G standalone (5G SA) system.

Current issues in cloud security

Keywords: cloud, artificial intelligence, supply chain, cybersecurity

This paper highlights the inevitability of cloud systems, provides examples of security issues that are specific to clouds, and then presents a possible methodology for deciding whether a cloud system meets the minimum required security level of the organization. From a security perspective, it outlines the key differences between a governmental and a public cloud, and illustrates the importance of supply chain protection through examples. It provides examples of the usage of artificial intelligence and its cloud-based solutions for defense purposes, as well as its usage on the offensive side.

Actual security challenges of cloud infrastructure from telecom perspective

Keywords: cloud infrastructure, open source, CVE, security, vulnerability

Nowadays, the infrastructure and platform variants used by telecom applications are highly IT based solutions. This is mostly beneficial from technical and business perspective, but also making the telecom systems more vulnerable against classical IT-specific attack methods. Telecom applications are considered as business critical systems, therefore it is very important to follow how these classical IT threats are effecting the telecom platforms and how the telecom systems can be efficiently protected against these new threats. A specific area dramatically increasing the threat level is the usage of OSS components, built in on all layers of telecom solutions. This paper is also covering non-technical aspects, like the evolution of security perception applied for telecom solutions.

The role of new codecs and AI in video processing evolution

Keywords: video encoding, machine learning, motion estimation, AI-based compression technology

Since economic and environmental aspects often place strict constraints on resources, encoder design traditionally relied on codec experts to develop heuristics and algorithms to shortlist the encoding tools for each application, according to pre-defined encoding efficiency or computational footprint targets. Although this approach may provide predictable and consistent reductions on processing requirements, understanding how each tool impacts compression efficiency for each content type and how they interact with each other makes it hard to scale, especially since the relationship between encoder complexity and compression efficiency is distinctly non-linear. MediaKind has been leveraging the use of ML to perform this analysis, powering real-time AI driven encoder decisions which proved to be far more efficient than any human defined heuristic or algorithm.

Almanach 2021

A special section of this issue contains one-page abstracts written by winners of HTE awards for best diploma theses in 2021.