Integrált LEO műholdas és földi hálózatok QoS jellemzőinek predikciója autoenkóder alapú gépi tanulással

Prediction of QoS characteristics of integrated LEO satellite and terrestrial networks using autoencoder-based machine learning

Dr. GÁL Zoltán

Debreceni Egyetem Informatikai Kar
4028 Debrecen, Kassai út 26., Magyarország, gal.zoltan@inf.unideb.hu, http://www.inf.unideb.hu

Abstract

We analyzed over 1 million RTT samples from Starlink's LEO system collected over 30 days, revealing significant fine-grained temporal variability. The RTT distribution shows dual peaks, indicating dynamic link-layer modes. Hourly aggregated features capture behavioral trends, while an autoencoder-based imputation reconstructs missing data effectively. We also evaluate RTT characteristics under VPN usage, highlighting performance shifts. These insights inform good modelling and robust interpretation of RTT behavior in NTN-TN integrated networks.

Keywords: Starlink, Non-Terrestrial Network, Autoencoder, Low Earth Orbit Satellites, Round-Trip Time, Machine Learning.

Kivonat

A Starlink LEO rendszeréből 30 nap alatt gyűjtött több mint 1 millió RTT-mintát elemeztünk, amely során jelentős, finomszemcsés időbeli változékonyságot tártunk fel. Az RTT-eloszlás kettős csúcsokat mutat, ami dinamikus adatkapcsolati üzemmódokat sugall a szolgáltatásra vonatkozóan. Az óránkénti összesített jellemzők rögzítik a viselkedési trendeket, míg az autoenkóder hatékonyan rekonstruálja a hiányzó adatokat. VPN-használat esetén is értékeljük az RTT-jellemzőket, kiemelve a teljesítménybeli eltolódásokat. Ezek a jellemző metrikák megfelelő modellezést és az RTT viselkedésének hatékony értelmezést teszik lehetővé NTN-TN integrált hálózatokban.

Kulcsszavak: Starlink, nem Földi hálózat, autoenkóder, alacsony Föld körüli pályán keringő műholdak, körforgási idő, gépi tanulás.

1. BEVEZETÉS

A műhold alapú nem földi hálózati (NTN – Non Terrestrial Network) szolgáltatásra való növekvő támaszkodással az elmúlt öt évben elindították a SpaceX által üzemeltetett műholdas internet-konstellációt, az úgynevezett Starlinket. Ez jelentős javulást jelent az alacsony Föld körüli pályán (LEO – Low Earth Orbit) keringő műholdakon keresztüli internetkapcsolat biztosításában, valamint a rosszul lefedett régiók lefedettségében. A LEO műholdas kommunikációs kapcsolatok dinamikus jellege azonban a földfelszíni hálózatoknál (TN – Terrestrial Network) kisebb változékonyságot eredményez a hálózati teljesítményben, például a késleltetésben, a dzsitterben és az átviteli sebességben, ami által kritikus fontosságúvá váltak a valós idejű alkalmazások felhasználói élményének minősége szempontjából.

A körforgási idő (RTT – Round Trip Time) egy széles körben használt mérőszám a hálózati kommunikáció késleltetésének felmérésére, amely a csomag forrástól a célállomásig és visszajutásának idejét jelenti. Míg a korábbi tanulmányok a műholdas hálózatok átlagos teljesítményét értékelték [1][2], az időbeli mintákat, a változékonyságot és a magasabb rendű statisztikai jellemzőket vizsgáló átfogó elemzések továbbra is korlátozottak, különösen a felhasználó oldali NTN-telepítések esetében.

Ebben a dolgozatban egy elemzést mutatunk be, amely Starlink antennáról 30 egymást követő napon mintavételezett RTT adatokon alapul, a TN hálózaton lévő nyolc különböző IP csomópont folyamatos pingelésével. A mérés során perc léptékű RTT értékeket rögzítettünk, ami lehetővé tette számunkra, hogy

minden IP-címhez két időskála szerinti elemzési mátrixokat hozzunk létre, amelyek sorai perceket, az oszlopok pedig napokat jelölnek. Ezek a mátrixok nagy felbontású képet adnak a hálózati teljesítmény dinamikájáról mind napi, mind hosszabb szintű időskálákon.

A valós LEO műholdas hálózati mérések, a fejlett jellemzőmérnökség és az időbeli elemzés összekapcsolásával a tanulmány célja, hogy hozzájáruljon a TNT és NT heterogén összekapcsolt internetes rendszerek teljesítmény-dinamikájának megértéséhez és méréséhez.

A dolgozat második részében a területen végzett kapcsolódó munkákat tárgyaljuk. Mivel a Starlink rendszer belső hálózati architektúrájáról korlátozott mennyiségű nyilvános információ áll rendelkezésre, a harmadik részben egy módszertant javasolunk a heterogén NTN és TN hálózatok RTT-metrikáinak mintavételezésére, mérésére és vizualizálására. A negyedik részben egy autoenkóder módszer eredményeit ismertetjük a hiányzó jellemző adatok pótlására, majd a végén összefoglaljuk a lényeget és következtetéseket vonunk le.

2. IRODALMI ÁTTEKINTÉS

2.1. Starlink műhold teljesítménye és jelelemzés

A legújabb kutatások széles körben vizsgálták a Starlink LEO műholdas hálózatának teljesítményét és környezeti hatásait. Ezek a vizsgálatok értékes betekintést nyújtanak a rendszer késleltetési viselkedésébe, a jelszerkezetekbe és a műholdas internetes szolgáltatásokkal kapcsolatos kihívásokba. A [3]-ban szereplő szerzők egy szimulációs platformot fejlesztettek ki a Starlink Ku-sávú letöltéséhez, amely a kulcsfontosságú jelszerkezeteket és a csatornahatásokat modellezi. Ez a munka jól segít az RTT mérésekhez kapcsolódó jelviselkedés megértéséhez. Egy másik tanulmányban [4] a kutatók átfogó elemzést nyújtanak a LEO műholdas kommunikációs dinamikájáról, a szolgáltatás minőségére, a késleltetésre és a Doppler-eltolódásra összpontosítva. Ezáltal alapvető betekintést nyújt a műholdas internetes rendszerek időbeli tulajdonságaiba. További szerzők elemzést végeztek a Starlink teljesítményéről egy jelentős napvihar idején, betekintést nyújtva a LEO műholdas hálózatok ellenálló képességébe és a kihívásokba, amelyekkel a szélsőséges űridőjárási körülmények között szembesülnek [5]. A [6]-ban a Starlink közép-európai mobil teljesítményéről készült méréseket mutattak be, az RTT-re, az átviteli sebességre és a csomagvesztésre összpontosítva a járművek mozgása során. Mindazonáltal egy másik értékelés is készült a Starlink hálózati teljesítményéről, amely a globális internetkapcsolatra gyakorolt hatásait tárgyalja [7].

2.2. Jellemző tulajdonság vektor kinyerése és RTT-elemzés a hálózati teljesítményben

A körbejárási idő kontextusában egy tanulmány azt vizsgálja, hogy a rendelkezésre álló monitorozási metrikák hogyan korrelálnak az RTT-ingadozásokkal, és irányokat javasol a teljesítményoptimalizálásra [8]. A [9]-ben bemutattak egy SuperFE nevű jellemzőkivonási technikát a forgalmi jellemzők hatékony és rugalmas kinyerésére gépi tanulási alkalmazásokhoz. Az értelmes jellemzők kinyerése az RTT-adatokból kulcsfontosságú a hálózati teljesítmény megértéséhez és a prediktív modellek fejlesztéséhez. A legújabb szakirodalom a jellemzőkinyerés különböző technikáira és azok hálózati elemzésben való alkalmazására összpontosított. Ezenkívül a [10]-ben egy másik jellemzőkinyerési technikát javasoltak, amelyet Packet2Vec-nek neveznek Word2Vec használatával, és amelyet relevánsnak tartanak a hálózati teljesítményelemzés szempontjából. A [11] dolgozat szerzői olyan felmérést készítettek, amely áttekinti a gépi tanulási megközelítéseket a teljesítmény javítása érdekében a vezeték nélküli hálózatok különböző rétegein, beleértve a jellemzőkinyerési technikákat is. Ez másik munka során az IoT-hálózat elemzését egy behatolásérzékelési jellemzőkinyerési technikával végezték, amely alkalmazható az RTT-elemzésben is [12].

2.3. Autoenkóder alapú adatgenerálás

Az autoenkóderek, különösen a variációs autoenkóderek (VAE) és a zajszűréses autoenkóderek (DAE), hatékony alkalmazhatóságukat bizonyították a hiányzó adatok kezelésében és hasonló adathalmazok generálásában. Az adatok rejtett (látens) reprezentációinak megértési képessége alkalmassá teszi ezeket a hiányos bemenetek rekonstruálására és a valósághű adatminták szimulálására. A legújabb tanulmányok az autoenkóderek hatékonyságát vizsgálták különböző területeken. A szintetikus adatgeneráláshoz használt variációs autoenkóderek kontextusában a [13]-ban a szerzők mélyreható áttekintést nyújtanak a VAE-kről, kiemelve azok képességét a kiváló minőségű szintetikus adatok generálására, miközben megőrzik a valós adathalmazok statisztikai tulajdonságait. Ezen kívül a DAE autoenkóder modell erős előállítási képességet

mutatott a hiányzó adatok széles skáláján, hangsúlyozva robusztusságukat és alkalmazkodó képességüket [14]. A kettős korrupciós DAE (DC-DAE) fokozta az általánosítást és megakadályozta a túlillesztést, ami a [15]-ben jobb pontosságot eredményezett.

3. ALKALMAZOTT MÓDSZERTAN

Ebben a dolgozatban az egyik széles körben használt műholdas hálózatot, a LEO Starlinket tanulmányozzuk és elemezzük. A Starlink egy nagyméretű LEO műhold konstelláció, amelyet a SpaceX telepített a globális szélessávú internet-hozzáférés biztosítására. A hagyományos geostacionárius műholdakkal ellentétben, amelyek körülbelül 35 786 km-es magasságban keringenek, a Starlink műholdak 340 km és 600 km közötti magasságban működnek. Ez az alacsonyabb pályamagasság jelentősen csökkenti az oda-vissza késleltetést, így alkalmassá teszi a késleltetésre érzékeny kommunikációs alkalmazásokhoz, miközben lehetővé teszi a gyors műholdas kapcsolatok roaming-ját és a hálózati teljesítmény fokozott tér-időbeli változékonyságát. A LEO rendszerek dinamikus viselkedésük miatt kihívásokkal szembesülnek. Ilyenek a Doppler-eltolódás, időben változó kapcsolatminőség, valamint a műholdak és a földi állomások közötti gyakori átmenetek. E hálózatok finomszemcsés időbeli viselkedésének megértése folyamatos monitorozást és munkamenet szintű elemzést igényel.

3.1. Adatgyűjtési keretrendszer

A Starlink LEO műholdas hálózatokon keresztüli RTT viselkedésének elemzéséhez létrehoztunk egy adatgyűjtési keretrendszert, amely az Internet Control Message Protocol (ICMP) visszhang kéréseken (azaz ping) alapul. Ezeket egy Starlink antennán keresztül forgalmaztuk. A nyolc különböző IP-cím közül öt nyilvános (PUB-bal jelölve) és három privát IP-címmel (PRI-vel jelölve) rendelkezett. Az RTT-méréseket körülbelül 1,5 másodpercenként végeztük el, ami elegendő az ICMP csomagok válaszidejének mérésére. Bizonyos technikai és szolgáltatási megfontolások miatt az RTT-mérések leálltak, ami hiányzó intervallumokat okozott a teljes mérési időszak alatt. Az RTT-mérésünk 2025. március 11-én kezdődött és 2025. május 5-én ért véget, de ebből csak egy 30 napos részt elemeztünk. Ezalatt IP-címenként több mint egymillió mintavételt kaptunk.

3.2. Előfeldolgozás és vizualizáció

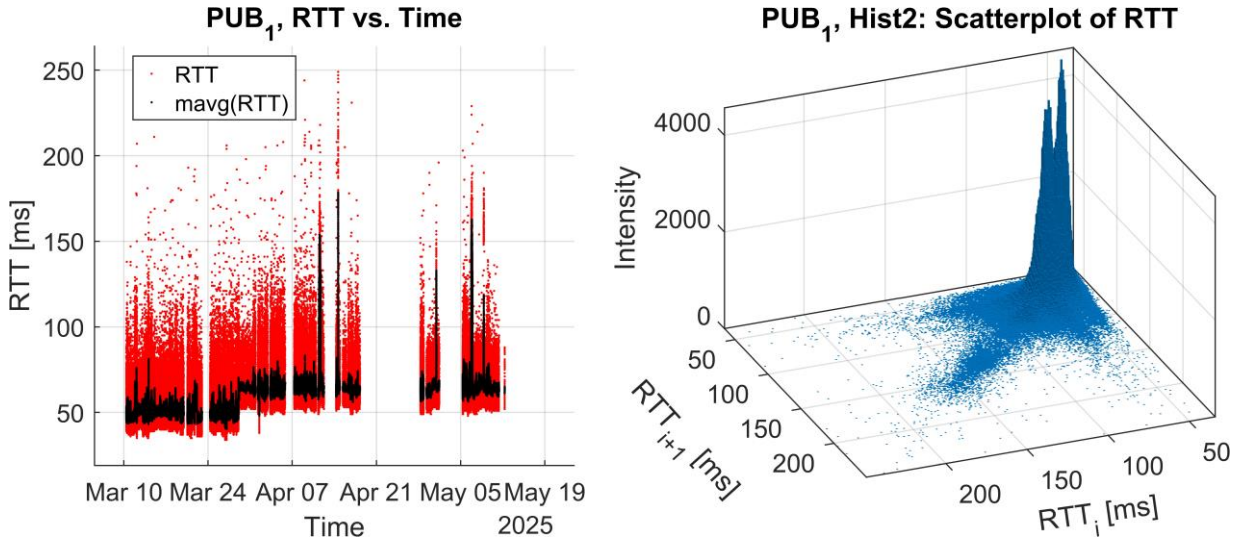
A mérési forgatókönyvben a LEO előfizetői számítógép egymást követő ping-műveleteket hajt végre nyolc különböző IP-cím célállomásával. Az adott intézmény első öt IP-címe nyilvános (PUB1, PUB2, …, PUB5), az utolsó három célhely pedig privát cím (PRI6, ..., PRI8). A nyilvános célcsomópontok akkor érhetők el, amikor a LEO előfizető végrehajtja a mérést. A privát célcsomópontokhoz VPN-en (Virtual Private Network) keresztül történt a hozzáférés, a VPN-munkamenet aktív állapota alatt.

Ping-elt csomópontok funkciója 1. táblázat

IP cím	Típus	Csomópont funkciója
PUB1	Public	MAN Domain Name Server
PUB2	Public	MAN Core Level Router
PUB3	Public	MAN VPN Server (Cisco Secure)
PUB4	Public	MAN Core Level Router
PUB5	Public	MAN Core Level Router
PRI6	Private	LAN Core Level Router
PRI7	Private	LAN Campus Level Router
PRI8	Private	LAN Campus Level Router

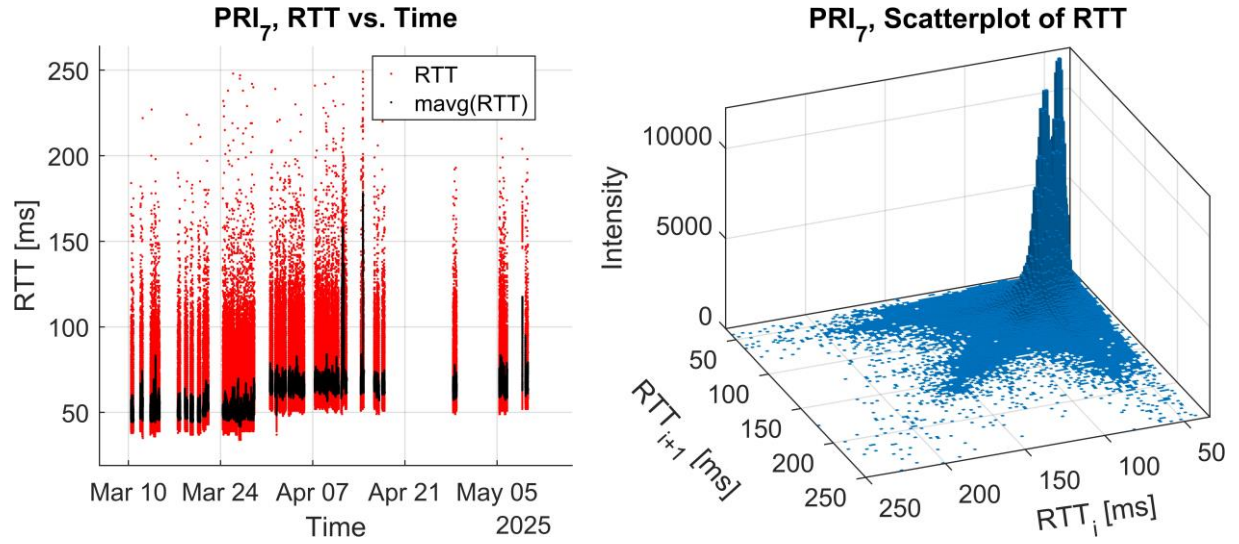
A célhálózat VPN-kiszolgálója általában naponta véletlenszerűen végrehajtja a munkamenetek lezárását a saját hardver erőforrásai megőrzése érdekében, mivel csak korlátos darabszámú egyidejű VPN kapcsolat működtetésre képes. A LEO előfizetői gép egymást követő mintavételezéseit az előző ping válasz után

(körülbelül 1,5 másodperc) hajtotta végre. A mérési időintervallum második felében a mintavételek ritkasága miatt (lásd az 1. és 2. ábrát) az elemzési időintervallumot az első 30 napra csökkentettük.



1. ábra. A nyilvános IP1 csomópont oda-vissza idő jellemzői (RTT értékek és (az RTT szóródási diagram 3D hisztogramja).

Ebben az elemzési időszakban 2025. április 27. körül 10 ms-os növekedést figyelhetünk meg az RTT mozgóátlag értékeiben. Ezenkívül a privát IP-címek mintavételezése ritkább a VPN-kiszolgáló hardver erőforrás-takarékos tevékenysége miatt. Mindkét 3D hisztogram kettős kiválasztást mutat, ami bizonyítja a szolgáltatási mód változását a heterogén NTN és TN hálózatokban (1. és 2. ábra). Az RTT válaszértékek (50, 250) ms tartományban vannak, ami jól mérhető az 1,5 másodperces időközökben.



2. ábra. A privát IP7 cím körfogási idő jellemzői (RTT értékek és az RTT szóródási diagram 3D hisztogramja).

Az ICMP csomagok kiesése esetén megjelenő, 250 ms-nál nagyobb értékeket kizártuk a grafikonokról. A hisztogramok 3D aspektusa javítja az RTT értékek egymást követő sorrendjének láthatóságát.

3.3. Munkamenet-mátrix felépítése és a jellemző vektorok kinyerése

A mélyebb időbeli elemzés érdekében minden IP-címhez kétdimenziós munkamenet-mátrixot képeztünk. A két munkamenet időtartama egy óra, illetve egy nap. Így a mátrix dimenziói a következők voltak: sorok: órák száma egy napban (24 óra); oszlopok: napok (csak az első 30 napot vesszük figyelembe). A mátrix munkamenet minden M_{ij}sorozata a j-edik nap i-edik órában mért RTT értékeket jelenti. Meg kell említeni, hogy a hiányzó intervallumok miatt nem minden M_{ij} munkamenet tartalmaz adatot. Emiatt a munkameneteket

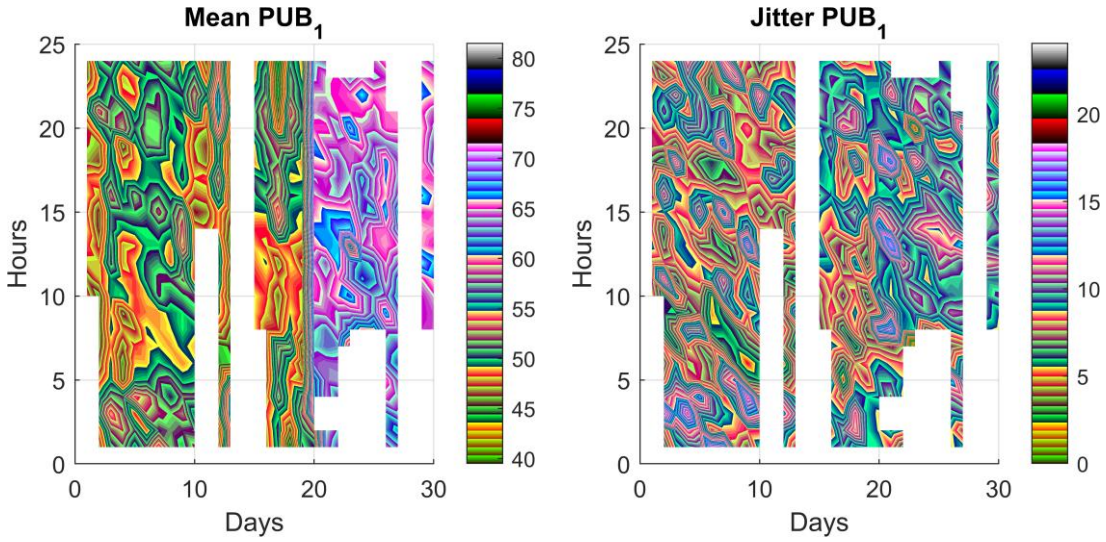
három csoportba soroltuk: E (Empty) - üres munkamenet, ami azt jelzi, hogy ebben az időintervallumban nem mértek RTT-ket; P (Partial) - részleges munkamenet, ahol a munkamenet tartalmaz RTT-ket, de nem a teljes intervallumra vonatkozóan; C (Complete) - teljes munkamenet esetén a munkamenet a megfelelő RTT-kkel rendelkezik a teljes időintervallumra vonatkozóan.

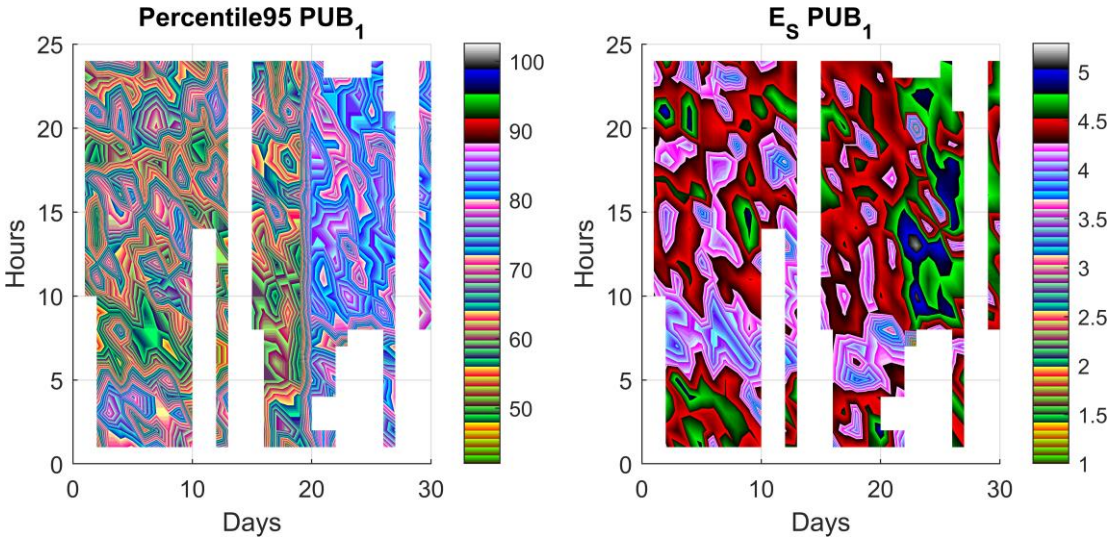
Minden M_{ij} munkamenetből egy 10 dimenziós jellemzővektort nyertünk ki, amely az RTT idősorok kapcsolódó statisztikai tulajdonságait rögzíti. A 2. táblázat a jellemzők kinyeréshez használt statisztikai metrikákat mutatja.

A munkamenetenként kinyert RTT-metrikák 2. táblázat

Metrika	**Szim-bólum**	**Definíció és tulajdonság**
Átlag	μ	Átlagos RTT érték a munkamenet során. A hálózati késleltetés központi tendenciáját rögzíti.
RTT tartomány	R	A maximális és minimális RTT értékek közötti különbség. Kiemeli a volatilitást vagy a börsztösséget.
Szórás (STD)	σ	Az RTT értékek szóródása az átlag körül. Az RTT változékonyságát méri.
Dzsitter	J	Az egymást követő RTT értékek közötti átlagos abszolút különbség. Az időbeli ingadozást (instabilitást) számszerűsíti.
Ferdeség	S	Az RTT eloszlás aszimmetriájának mérése. A magasabb vagy alacsonyabb késleltetések felé mutató torzítást azonosítja.
Hurst paraméter	H	A hosszú távú függőség becslése (0,5–1). Az RTT mintákban memóriahatásokat és önhasonlóságot észlel.
95 percentilis	P_{95}	Az RTT érték, amely alá a munkamenet minták 95%-a esik. A legrosszabb esethez közeli késleltetést jelzi.
99 percentilis	P_{99}	Az RTT érték, amely alá a munkamenet minták 99%-a esik. A késleltetés farok viselkedését rögzíti.
Shannon Entrópia	E_s	Az RTT hisztogram entrópiája (információtartalom). Az RTT értékek bizonytalanságát és diverzitását méri.
Spectrális Entrópia	E_f	A teljesítményspektrum-sűrűség eloszlás entrópiája (FFT-n keresztül). A frekvenciatartománybeli szabálytalanságokat és komplexitást azonosítja.

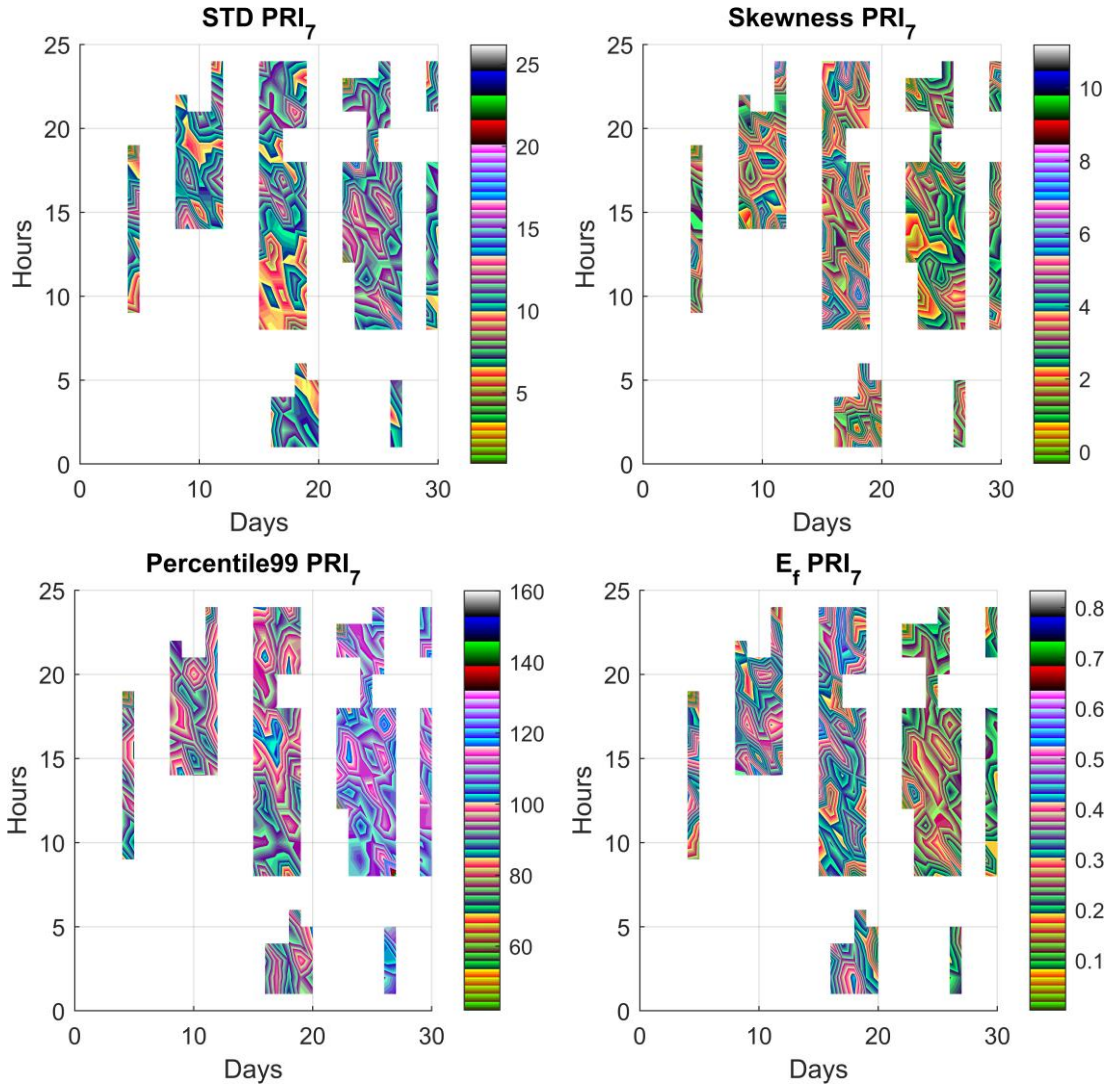
Ezeket a jellemzőket minden IP-címre külön-külön számítottuk ki. A jellemzők többsége a heterogén NTN és TN hálózatok kiszolgálási módját érzékeli mind a PUB, mind a PRI címek esetében (lásd a 3. és 4. ábrák).





3. ábra. A PUB1 IP-csomópont jellemzővektor-metrikái (átlag, dzsitter, 95-ös percentilis és Shannon-entrópia).

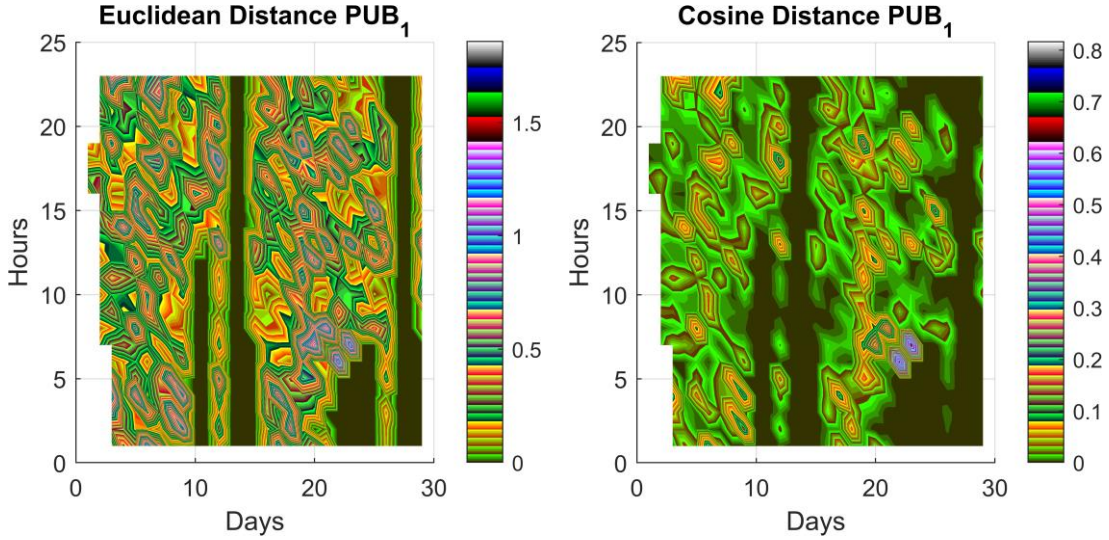
Az elemzési időszakban a nyilvános és privát IP-címek válaszintenzitása nagyon eltérő a VPN-kiszolgáló tulajdonság miatt.



4. ábra. A PRI7 IP-csomópont jellemzővektor-metrikái (STD: szórás, ferdeség, 99-es percentilis és spektrális entrópia).

Miután kiszámítottuk az egyes munkamenetek jellemző vektorait, két kiegészítő távolságmetrika segítségével mértük a munkamenetek közötti hasonlóságot: euklideszi és koszinusz távolság. Az euklideszi távolság két jellemzővektor közötti abszolút geometriai különbséget mutatja meg, betekintést nyújtva azok nagyságrenden alapuló különbségébe. Ezzel szemben a koszinusz-távolság a vektorok közötti szögkülönbséget értékeli, hangsúlyozva azok irányultságát a skálától függetlenül. A koszinusz-távolság metrika ebben az esetben 1-cos(α), ahol α a két 10 dimenziós jellemzővektor közötti szög. Ily módon a koszinusz-metrika kis értékű az α kis szög esetén.

E két távolság feldolgozása jellemzővektorok között történik, amelyek az egymást követő órákban és napokon lévő szomszédos munkameneteknek felelnek meg (lásd az 5. ábrát).

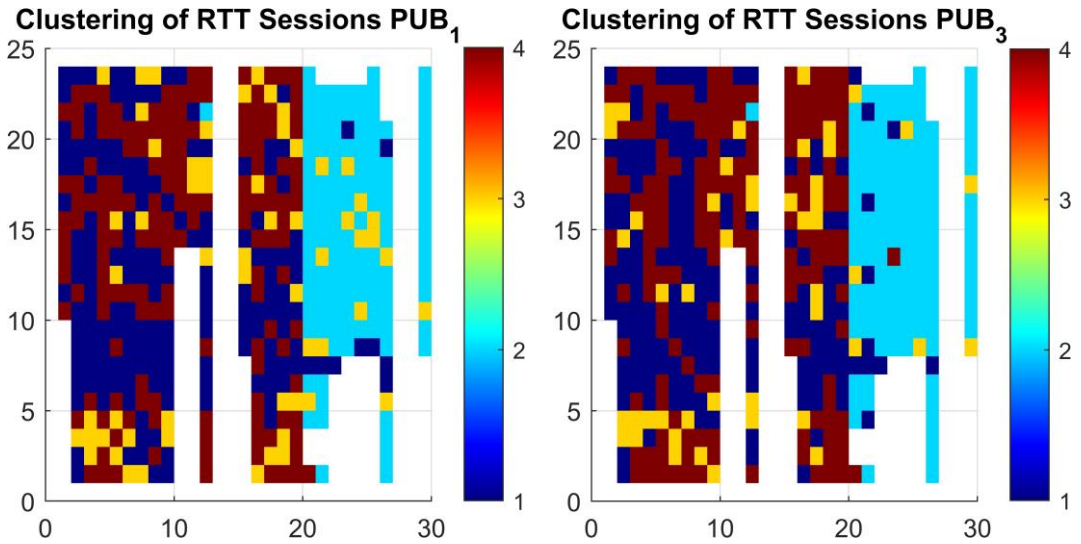


5. ábra. A munkamenetre normalizált jellemzővektorok közötti távolság a PUB1 IP-csomópont vízszintes és függőleges irányában (euklideszi és koszinusz távolságok).

A távolságábrák az egymást követő munkamenetek fraktál tulajdonságát mutatják, amelyet a 10 dimenziós jellemzővektorok jellemeznek (lásd a 3., 4. és 5. ábrát). A távolságábrák homogén területeit az üres munkamenetek rései okozzák. A szomszédos jellemzővektorok közötti euklideszi távolság nagyobb változékonyságot mutat, mint a koszinusz távolság ugyanazon munkamenet-pár esetében (lásd az 5. ábrát).

4. IV. GÉPI TANULÁS FOLYAMATOKHOZ, FUNKCIÓK KÉPZÉSE

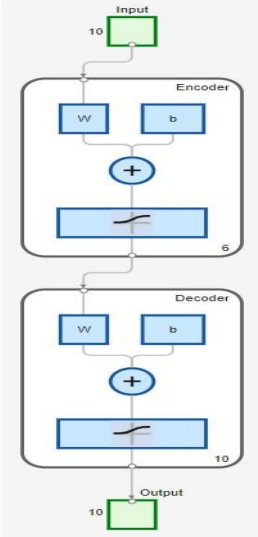
A rekonstruált munkamenet-adatok mögöttes mintázatainak feltárásához a K-Mean klaszterezési algoritmust alkalmaztuk K=4 értékkel.



6. ábra. Az RTT-munkamenetek klaszterezése (PUB1 és PUB3 IP-csomópontok).

Ez a felügyelet nélküli tanulási módszer a jellemzővektorokat négy különálló csoportra osztja a klasztereken belüli variancia minimalizálásával és a klaszterek közötti távolság maximalizálásával. A négy klaszter kiválasztását empirikusan határoztuk meg az értelmezhetőség és a mintafelbontás egyensúlyának megteremtése érdekében. Minden klaszter hasonló RTT-dinamikával rendelkező munkameneteket rögzít, lehetővé téve számunkra, hogy megkülönböztessük az alacsony késleltetésű stabil viselkedéseket, a magas dzsitter intervallumokat, a részleges zavarokat és az anomáliás mintákat (lásd a 6. ábrát). Ez a klaszterezés elősegíti az időbeli teljesítménytartományok mélyebb megértését a megfigyelt IP-címek között. A munkamenetek klaszterezési eredménye megfelel az IP-cím típusának (PUB, PRI). A nyilvános címek megfelelő munkamenetei között erős korreláció létezik. Hasonló viselkedés figyelhető meg a privát címekhez tartozó munkamenetek esetében is.

A tanulmányban egy autoenkóder modellt alkalmaztunk a technikai zavarok okozta hiányos munkamenet-adatok problémájának kezelésére. Az autoenkóderek olyan neurális hálózatok, amelyek célja az adatok hatékony reprezentációinak megtanulása a bemenet látens térbe történő tömörítésével, majd annak rekonstruálásával.



Az autoenkóder definíciója:

Bemenet: X_norm;

Paraméterek:

Hiddensize = 16;

MaxEpochs = 100;

L2WeightRegularization = 0.001;

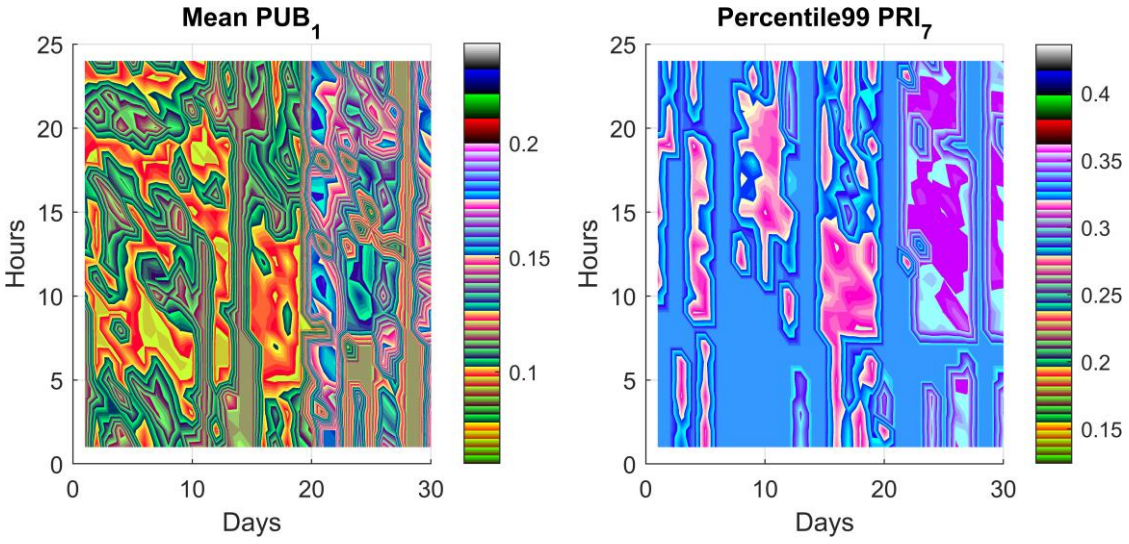
SparsityRegularization = 4;

SparsityProportion = 0.05;

ScaleData = True;

7. ábra. A használt autoenkóder modellje és jellemzői. Balra: autoenkóder modell; Jobbra: az autoenkóder definíciója.

A modell teljes munkameneteken történő betanításával az autoenkóder megtanulja az RTT jellemzővektorok belső szerkezetét, lehetővé téve számára, hogy valószínűsíthető rekonstrukciókat generáljon a hiányzó vagy részben megfigyelt munkamenetekhez.



8. ábra. A PUB1 (átlag) és a PRI7 (99. percentilis) IP-csomópontok autoenkóder alapú kitöltött jellemzővektor metrikái.

Ez a megközelítés nemcsak az adatok időbeli és statisztikai koherenciáját őrzi meg, hanem a későbbi elemzés és modellezés megbízhatóságát is növeli. Az autoenkóder kitöltési folyamatának minőségi eredményei az üres szomszédos munkamenetek számától függenek. Minél nagyobb az üres szomszédos munkamenetek száma, annál homogénebb kitöltési választ hajt végre az autoenkóder (lásd a 8. ábrát).

A jellemzővektor dimenzióiban lévő különböző metrikák eltérő érzékenységgel rendelkeznek az RTT-értékek változási frekvenciáira. Az alacsony frekvenciájú események aluláteresztő szűrő jellegű metrikák, például az Átlag és a Tartomány metrika segítségével detektálhatók. Ezért a 8. ábrán megfigyelhető a heterogén NTN és TN hálózatok két üzemmódja, a 19. napon bekövetkező változással. A nagy frekvenciájú eseményeket felüláteresztő szűrő metrikák, például a 99. percentilis és a dzistter detektálják.

Az autoenkóder által készített kitöltési folyamat mennyiségi jellemzéséhez külön-külön kiszámítottuk az eredeti és a kitöltött munkamenet-távolság mátrixok négyzetes középértékének hibáját (RMSE – Root Mean Square Error) euklideszi és koszinusz távolságokkal (ld. 3. táblázat). Az eredeti és az autoenkóderrel kitöltött munkamenetek jellemző vektorai közötti RMSE szemléletesen érzékelteti az Euklideszi illetve Koszinusz távolság metrikák képességét a publikus és a privát IP című csomópontok válaszidejére vonatkozóan.

Az eredeti és az autoenkóderrel kitöltött jellemző vektorok RMSE hasonlósága. *3. táblázat*

RMSE tartománya	PUB IP című csomópontok	PRI IP című csomópontok
Euklideszi távolság	0.66 …0.72	0.58 … 0.64
Koszinusz távolság	0.17 … 0.22	0.18 … 0.23

A fő különbség a célcsomópontok típusában van. A PUB IP-csomópontok euklideszi távolságának tartománya nagyobb, mint a PRI IP-csomópontok tartománya. Ennek oka a teljes munkamenetek nagyobb száma az autoenkóderrel kitöltött munkamenetek után. Másrészt azt tapasztaljuk, hogy a koszinusz távolságok esetén az RMSE tartomány gyakorlatilag független a célcímek típusától.

5. ÖSSZEFOGLALÁS

Ez a tanulmány a Starlink LEO műhold RTT-adatainak nagy felbontású időbeli elemzését mutatja be, kiemelve a változékonyságot mind a napi, mind a hosszú távú skálákon. Több mint 1 millió ping-minta 30 napon keresztül történő felhasználásával a tanulmány jelentős RTT-ingadozásokat tár fel, amelyeket a műholdak dinamikája és a VPN-munkamenet viselkedése befolyásol. Az óránkénti munkamenet-mátrixokból származó jellemzők kinyerése időbeli mintázatokat azonosított be, beleértve a kettős RTT-csúcsokat, amelyek a szolgáltatási mód változásait jelzik. Egy autoenkóder hatékonyan generálta a hiányzó adatokat, megerősítve a módszer hasznosságát a valós NTN-TN rendszerekben. A részletesebb elemzéseket, mint például a Markov-lánc Monte Carlo-analízisét, a TensorFlow valószínűségét, az RTT és a Starlink műhold pályaparamétereinek korrelációját, az autoenkóder rekonstrukciós hibáinak jellemzőtípus szerinti elemzését és a klaszter stabilitását RMSE és MAE metrikák alapján, a kutatási munka kibővített változatában értékeljük.

KÖSZÖNETNYILVÁNÍTÁS

Ezt a munkát a Debreceni Egyetem QoS-HPC-IoT Laboratórium és a TKP2021-NKTA-34 projekt támogatta. A TKP2021-NKTA-34 számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a TKP2021-NKTA pályázati program finanszírozásában valósult meg.

IRODALMI HIVATKOZÁSOK

[1] Laniewski, Dominic, et al. *Wetlinks: a large-scale longitudinal starlink dataset with contiguous weather data.* 2024 8th Network Traffic Measurement and Analysis Conference (TMA). IEEE, 2024.

[2] Mohan, Nitinder, et al. *A multifaceted look at starlink performance.* Proceedings of the ACM Web Conference 2024. 2024.

[3] Komodromos, Zacharias M., Wenkai Qin, and Todd E. Humphreys. *Signal simulator for Starlink Ku-Band downlink.* Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023). 2023.

[4] Gal, Zoltan, and Djamila Talbi. *Insights into low Earth orbit satellite communication dynamics: quality of service analysis of connection behavior, latency, and Doppler shift.* 2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB). IEEE, 2024.
[5] Ramanathan, Alagappan, and Sangeetha Abdu Jyothi. *Weathering a Solar Superstorm: Starlink Performance during the May 2024 Storm.* Proceedings of the 2nd International Workshop on LEO Networking and Communication. 2024.
[6] Laniewski, Dominic, et al. *Starlink on the Road: A First Look at Mobile Starlink Performance in Central Europe.* 2024 8th Network Traffic Measurement and Analysis Conference (TMA). IEEE, 2024.
[7] Wong, Aaron, et al. *Network Performance Analysis of Starlink: A Proliferated Low Earth Orbit Satellite Internet System.* 2024 17th International Conference on Signal Processing and Communication System (ICSPCS). IEEE, 2024.
[8] Giannakopoulos, Panagiotis, et al. *perfCorrelate: Performance variability correlation framework.* Future Generation Computer Systems 170 (2025): 107827.
[9] Zhang, Menghao, et al. *SuperFE: A Scalable and Flexible Feature Extractor for ML-based Traffic Analysis Applications.* Proceedings of the Twentieth European Conference on Computer Systems. 2025.
[10] Goodman, Eric L., Chase Zimmerman, and Corey Hudson. *Packet2vec: Utilizing word2vec for feature extraction in packet data.* arXiv preprint arXiv:2004.14477 (2020).
[11] Kulin, Merima, et al. *A survey on machine learning-based performance improvement of wireless networks: PHY, MAC and network layer.* Electronics 10.3 (2021): 318.
[12] Sarhan, Mohanad, et al. *Feature extraction for machine learning-based intrusion detection in IoT networks.* Digital Communications and Networks 10.1 (2024): 205-216.
[13] Olaoye, Godwin, and Ayuns Luz. *Hybrid Models for Medical Data Analysis.* Available at SSRN 4742530 (2024).
[14] Abiri, Najmeh, et al. *Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems.* Neurocomputing 365 (2019): 137-146.
[15] Jiao, Xu, et al. *Improving Generalization for Missing Data Imputation via Dual Corruption Denoising Autoencoders.* International Conference on Applied Computational Intelligence, Informatics and Big Data. Springer, Cham, 2025.