

# Opponensi vélemény

Toka László

## ”Erőforrás-kezelés felhőrendszerekben”

c. DSc Értekezéséről

Az értekezés témaválasztása rendkívül korszerű. A földrajzilag kiterjedt, nagyméretű elosztott rendszerek, mint a peremfelhő és IoT rendszerek jelentősége folyamatosan nő és egyre több területen alkalmazzák őket. Ugyanakkor alkalmazásuk számos olyan problémát vet fel, amelyek megoldása jelentős kutatást igényel. Ilyen problémák a szolgáltatások optimális árazása, a feladatok optimális elosztása a peremfelhő számítási rendszerek csomópontjain úgy, hogy magas megbízhatóságot és alacsony késleltetést biztosítsanak. Szintén fontos gyakorlati probléma a szoftver rendszerek automatikus skálázása a használati intenzitásuk időbeli változásának figyelembevételével.

Az értekezésben alkalmazott kutatási módszertan megfelel a választott téma sajátosságainak. A nagyméretű elosztott rendszerek sok esetben olyan komplexek, hogy vagy egyszerűsítő modellek alkalmazása mellett lehetséges matematikai apparátussal kezelni őket, vagy még az egyszerűsítő modellek mellett sem kezelhetők matematikai apparátussal. Ahol a matematikai megközelítés lehetséges volt, ott a szerző igen helyesen ezt az utat választotta, mint például a 2. Fejezetben. Ahol ez nem volt lehetséges, mint pl. a 3.5 fejezetben, ott szimulációkra ill. valós adathalmazok feldolgozására támaszkodott. Ahol lehetséges volt, ott a matematikai modellel kapott előrejelzést a szerző összevetette a szimulációs eredményekkel és ezek meggyőzően verifikálták az alkalmazott szimulációs eszközt és módszert (ld. 3.5.1 fejezet). Mindez kellő hitellességet kölcsönzött azoknak a vizsgálatoknak, amelyeket már csak a szimulátoron tudott elvégezni a szerző.

### **Formai szempontok**

Az értekezés 5 fejezetből áll. Az első, bevezető fejezetben a problémakör felvázolása történik a szerző tudományos hozzájárulásának összefoglalásával és az alkalmazott módszertan, ill. a disszertáció struktúrájának bemutatásával. Az érdemi kidolgozás a 2., 3. és 4. fejezetben történik, amelyek a felállított 3 téziscsoport részletes kifejtését tartalmazzák. Ezen fejezeteket a szerző a kapcsolódó nemzetközi kutatások aktuális eredményeinek bemutatásával zárja, ezzel segítve az olvasót és a bírálót a jelölt önálló, új eredményeinek megítélésében. Az 5. fejezet áttekinti a 3 téziscsoportot, tételesen kimondva azokat a téziseket, amelyek a disszertációban leírt kutatások eredményeit foglalják össze. Hiányolok viszont egy végső összefoglaló és kitekintő fejezetet az értekezés végéről, ami egyrészt bemutatná az értekezésben elért tudományos eredmények gyakorlati hasznosulását, másrészt a kutatás jövőbeli folytatásának esetleges irányait. Hozzá kell tennem, hogy a téziszfüzet utolsó fejezete “Az eredmények hasznosíthatósága és hatása” címmel valamennyire pótolja ezt a

hiányosságot, de kívánatos lett volna ezt a részt az értekezésben is szerepeltetni, mivel a mű későbbi olvasói nem feltétlenül nézik meg / értik a magyar nyelvű téziszűzetet.

Az értekezés gondosan és szépen van szerkesztve. Az ábrák és a beszerkesztett képletek, ill. programrészletek jól olvashatók. Az értekezés nyelve angol, amit a szerző igen jó színvonalon használ. Nyelvtani hiba vagy elírás alig van a dolgozatban. A szakterület sajátosságának megfelelően a szövegben nagy számban található rövidítések, amelyek döntő többsége szerepel a szerző által összeállított magyarázó szójegyzékben (Glossary), de sajnos nem az összes.

A szakirodalmi hivatkozások száma 146. A jelölt által írt, vagy társszerzőként jegyzett publikációk száma 12, amiből 5 folyóirat cikk és 7 konferencia előadás/cikk. Sajnálatos, hogy semelyik hivatkozás nem tartalmaz oldalszámot, noha ezt szokás megadni, és a saját cikkek esetén mindenképpen fontos lett volna. A publikációkon kívül a jelölt 3 szabadalomban is társszerző, amelyek 2021-22-ben készültek. A hivatkozások száma elegendő annak megítélésére, hogy a szerző munkája hogyan illeszkedik a téma nemzetközi feldolgozottságába. Ennek alapján elmondható, hogy a szerző alaposan ismeri a téma nemzetközi szakirodalmát és annak fejlesztéséhez jelentős mértékben hozzájárult. Talán könnyebben kezelhető lenne a hivatkozáslista, ha névsor szerint lenne rendezve.

### **Az értekezés tartalmi áttekintése**

Az **1. Fejezet** egy jó összefoglaló a disszertációban leírt kutatás céljáról, a megoldandó problémakörökről és a kutatási módszerekről. Ezenkívül bemutatja a disszertáció felépítését és struktúráját.

A **2. Fejezetben** a jelölt modelleket és azokon végzett analitikai elemzéseket mutat be az 5G ökoszisztémában megjelenő közvetítői szolgáltatások szolgáltatók közötti árazásához, valamint a felhő- és hálózati szolgáltatások végfelhasználók számára történő árazásához Stackelberg-játékok alapján. A fejezetben leírt kutatás jelentősége, hogy kimutatta, hogy a felhő-központú szolgáltatásnyújtás hogyan befolyásolhatja az Internet szerkezetét, és mely tényezők befolyásolhatják a szolgáltatókat a szolgáltatások közvetítésének árazásában.

Mivel az ebben a fejezetben felépített matematikai modell több ponton a Stackelberg játékra támaszkodik, ajánlatos lett volna kicsit elidőzni ezen a témán a háttérodalom bemutatásakor, és pontosabban beazonosítani a modellben felhasznált variációkat, tekintettel arra, hogy az eredeti Stackelberg játéknak mára már számos módosított változatát alkalmazzák a legkülönbözőbb területeken.

Jó lenne, ha a disszertáció tartalmazna egy áttekintő listát a használt jelölések magyarázatával, ahogy az a mélyebb matematikai alapokon álló publikációknál bevált gyakorlat. Ennek hiányában az egyes képletekhez kapcsolódóan bevezetett definíciók/jelölések visszakeresése fáradságos és időigényes feladat az olvasó számára.

A jelölésekkel kapcsolatban néhány kisebb pongyolaság is található a dolgozatban. A 12. oldalon a 2.3. Lemmában szereplő "business matrix  $M$ " explicit formában nem lett korábban bevezetve, így időbe telik beazonosítani a 9. oldalon a 2.1. képletben bevezetett "extent of services"  $M_{ij}$  elemekkel, továbbá a 2.1. Lemmában szereplő "extent of business"-nek nevezett tényezőkkel. Az extent of services/ business fogalmát tartalmilag is szükséges lett volna alaposabban megmagyarázni, mert nem világos a mérhetősége.

A 22. oldalon a 2.10 lemmában a PoA képlet szöveges értelmezése hibás, mert az anarchia árához viszonyított aránynak van megnevezve (ratio to the the price of anarchy), holott valójában ez az anarchia ára (price of anarchy). Meg kell jegyezni, hogy a tézisfüzetben az állítás helyesen szerepel.

A 23. oldal alján levő számozatlan képletben, amely a 2.4c. ábrán látható esetre adja meg a játék értékét, a jelölések magyarázata nem precíz, összemossa a felhasználói kéréseket a hozzájuk kapcsolódó eseményekkel. A képletben szereplő  $\hat{c}_2$  (kalapos c) feltehetőleg egyszerű elírás, de tovább nehezíti a formula megértését.

Ugyanakkor elmondható, hogy a disszertációban szereplő képletek mennyiségéhez képest a fellelt kisebb pontatlanságok száma elenyésző. A jelölt nagyon komoly matematikai jártasságról tesz bizonyosságot az ebben a fejezetben felvonultatott matematikai apparátus mennyiségi és minőségi használatával. Szintén imponáló, hogy a modellezéshez alkalmazott eszközkészlet és elemzési módszertan a matematikai apparátus milyen széles skáláján mozog. Ezek felhasználásával és a felhő- és hálózati szolgáltatók 5G ökoszisztémájára való továbbfejlesztésével a jelölt olyan eredményeket állít elő, amit a jövőbeni online szolgáltatásokat nyújtó infrastruktúra-szolgáltatók szövetségei használhatnak fel szolgáltatásaik árazása során.

A 2. Fejezethez kapcsolódó téziscsoportban 5 tézis található, melyek mindegyikét elfogadom új tudományos eredményként a disszertációban bemutatott matematikai kifejtésük alapján.

**A 3. Fejezetben** három olyan problémát kíván megoldani a jelölt, amelyek a nagyméretű peremfelhő számítási rendszerekben jelentősen befolyásolják a rendszer minőségét és használhatóságát. Az első problémakör a késlelteteskritikus alkalmazások ütemezését kívánja úgy megoldani, hogy a rendszer megbízhatósági követelményeket is kielégítsen. Egy ilyen megbízhatósági követelmény előírja, hogy bármelyik csomópont meghibásodásából a rendszer hibamentes működését vissza lehessen állítani. Ennek érdekében a jelölt kidolgozott két perem-ütemező algoritmust, melyek lényege, hogy minden podhoz biztonsági mentési erőforrásokat u.n. helyőrzőket biztosít a peremfelhő csomópontjain.

A jelölt a fenti feltételeknek megfelelően kidolgozta mind az online perem-ütemező, mind az offline újraütemező algoritmusát. Helyesen mutat rá a disszertációban, hogy az ütemezési algoritmus futása közben bekövetkező tranzienis jelenségek problémákat okozhatnak. Idézem, hogy mit ír erről a disszertációban: "Ha ilyen események történnek az újraütemező futása közben, akkor az algoritmus által adott elhelyezési eredmény már nem biztos, hogy érvényes lesz." A jelölt által javasolt megoldás, hogy az algoritmusoknak kellően gyorsaknak kell lenni, azaz polinomiális időben kell lefutniuk és be is bizonyítja, hogy ezek az algoritmusok polinomiális időben futnak.

Valóban sokat segít, ha az algoritmus gyors, és polinomiális időben fut, de ez önmagában nem oldja meg a fenti tranzienis problémát, csak csökkenti az előfordulások számát. Sajnos a jelölt nem tárgyalja azt az esetet, amikor a fenti tranzienis probléma megjelenik, noha valós körülmények között ezt a problémát mindenképp kezelni kell és nem elegendő arra hagyatkozni, hogy az újraütemező algoritmus gyors. Mivel a tranzienis probléma nincs megoldva a disszertációban, ezért az ehhez a kutatáshoz kapcsolódó 2.1 tézist nem tudom elfogadni. A 2.1 tézis formálisan azt mondja ki, hogy az online ütemezés, az offline újraütemezés és az infrastruktúra csomópontszegmentálási módszerek polinomiális komplexitásúak. Ez valóban igaz állítás, a jelölt ad is rájuk bizonyítást, de ezek használhatóság nélkül nem jelentenek tézis értékű új tudományos eredményt.

A 2.2 tézisben a heurisztikus online ütemező algoritmus válaszában minőségét garantálja a jelölt, bizonyítva, hogy a javasolt online ütemezési módszer egy 3-approximációs algoritmus a Pod-ok helyőrzőinek együttes elhelyezésére. Ezt a tézist elfogadom új tudományos eredménynek.

A 3. Fejezet másik nagy témaköre a podok automatikus skálázása felhő rendszerekben. A kutatás kiindulópontja a gyakorlatban széles körben használt Kubernetes felhőkezelő platform, amely a Horizontal Pod Autoscaler (HPA) nevű beépített eljárással valósítja meg podok igény szerinti dinamikus skálázását. A kutatás alapkérdése itt az volt, hogy lehet-e a HPA algoritmusnál hatékonyabb algoritmust kidolgozni a podok skálázására. Ennek érdekében első lépésként a jelölt kidolgozta a HPA algoritmus matematikai modelljét melyben a kérések érkezési folyamatát Markov-modulált Poisson folyamatként (MMPP) írja le. Mivel ennek a modellnek hátránya, hogy nem tudja figyelembe venni az elveszett jobok számát, ezért a jelölt második lépésként egy olyan un. Diszkrét modellt vezetett be, ami már kiküszöböli ezt a problémát. A modellek kísérleti validálása érdekében a jelölt egy egyetemi hálózatról gyűjtött forgalmi adatokból összeállított valós működési mintákat felhasználva végzett összehasonlító szimulációs elemzéseket, melyek kimutatták, hogy a bevezetett matematikai modellek valóban jól írják le a HPA algoritmusát, azaz jó kiindulási alapot jelentenek a további vizsgálatokhoz.

Ezt a nagyon tartalmas és sokoldalú munkát foglalja össze a 2.3 Tézis meglehetősen egyszerűsített megfogalmazásban. A tézist nagyon komoly eredménynek tartom és elfogadom, de kár, hogy a jelölt nem fordított nagyobb energiát ennek a fontos tézisnek az informatívabb megfogalmazására.

A 2.4 Tézist nagyon színvonalas és sokrétű kísérleti munka támasztja alá, melynek keretében a jelölt 4 különböző gépi tanulási módszert (1/ autoregresszív (AR) modell, 2/ felügyelt mélytanulás módszer (LSTM), 3/ felügyelet nélküli mély tanulás módszer (HTM), 4/ megerősítő tanulás módszer) alkalmazott és vetett össze az általa kidolgozott diszkrét algoritmussal. A szimulációs kísérletek alapján egyik sem bizonyult egyértelműen kielégítőnek a HPA algoritmus feljavítására és ezért egy olyan modellt javasolt a szerző, ami ezeket a modelleket alkalmazza és dinamikusan váltogatja aszerint, hogy az előző időszakban melyik hogyan teljesített. Ezzel a módszerrel jelentős (esetenként akár 50%-os) költségmegtakarítást ért el a Kubernetes HPA algoritmusához képest. Ennek az eredménynek igen komoly gyakorlati alkalmazása is lehet, nem véletlen, hogy a jelölt szerzőtársaival együtt szabadalmat nyújtott be a témában. Mindezek alapján a 2.4 Tézist fontos eredménynek tartom és elfogadom.

A 2.5 Tézis foglalkozik a harmadik problémakörrel, a mikroszolgáltatások esetenkénti összevonásával elérhető teljesítmény növeléssel. Az ehhez kapcsolódó kutatásban a jelölt analitikai modellt javasolt a nagy léptékű alkalmazásfutási időszakok alatti erőforrás-többlet és az alkalmazás több skálázási egységbe rendezéséből fakadó késleltetési többlet közötti döntési helyzet leírására. Ezzel a tézissel kapcsolatban jelentős probléma, hogy a javasolt analitikai modell nem veszi figyelembe a skálázási egységek kommunikációs igényét. Ezt a szerző maga is megírja, de csak a modell matematikai bevezetése végén. Sokkal elegánsabb lett volna már az elején kimondani, hogy mik a javasolt modell korlátai, mert így a levezetések követésében kevésbé lett volna zavaró ennek a hiánynak az állandó megjelenése. Ennél is nagyobb probléma, hogy e korlát súlyosságát a jelölt nem vizsgálja, nem diszkutálja és így nem lehet tudni, hogy a valós problémák hány százalékát zárja ki ez a megszorítás. Ezt olyan komoly problémának tartom, ami az egész modell használhatóságát alapjaiban kérdőjelezi meg és ezért ezt a 2.5 Tézist nem fogadom el új tudományos eredménynek.

**A 4. Fejezetben** a jelölt két szolgáltatásminőség-biztosítási keretrendszert mutat be IoT rendszerekben a feltöltési sávszélesség optimális kiosztására. Az elsőben egy Vickrey típusú árverésen alapuló, koordinálatlan erőforrásallokációs modellt javasol a szerző felhőalkalmazások elérése során a hálózati QoS fenntartása érdekében. Itt a kliensek rövid időszakokra többlet sávszélességért licitálhatnak a hálózat üzemeltetőjénél történő ajánlattétellel, de a disszertációban nincs megindokolva, hogy a különböző árverési modellek közül miért épp a Vickrey típusú modell lett vizsgálva. A disszertáció ezen részében a jelölt numerikus elemzéssel olyan heurisztikus stratégiákat mutat be, amelyek bizonyos forgalmi terhelési esetekben sikerre vezetnek.

A második keretrendszerben egy központosított erőforrás-allokációs sémát javasolt a jelölt, ahol a felhőbe feldolgozásra küldött adatfolyamok sebességét egy, a felhőben futó optimalizáló szoftver képes előre jelezni és szabályozni. Ebben a modellben a megosztott feltöltési sáv szélesség az alkalmazás szempontjából fontos forgalmi adatok szerint van kiosztva. Az erőforrás allokációs problémát dinamikus programozással oldja meg a jelölt, és megadja a kidolgozott algoritmus számítási komplexitását. Ennek a megoldásnak egy konkrét gyakorlati feladat megoldásában is jelentős szerepe volt, és így komoly gyakorlati jelentőséggel bír.

A jelölt a két keretrendszerhez kapcsolódóan önálló téziseket mond ki, amelyeket külön-külön nem tartok tézisértékűen erős állításoknak. Ugyanakkor a két keretrendszer együttesen már kiad egy tézisértékű munkát és eredmény halmazt és ezért a 3.1 és 3.2 tézist összevonva, egy tézisként el tudom fogadni.

### **Összefoglalás**

Mindent összevetve az értekezés meggyőzően bizonyítja, hogy a szerző kiemelkedő szakértője a nagyméretű peremfelhő és IoT rendszereknek. Igen sok szempontból vizsgálja ezeket az új típusú elosztott rendszereket és sok területen jelentősen hozzájárult jobb megértésükhöz és mind elméleti, mind gyakorlati továbbfejlesztésükhöz.

A jelölt 3 téziscsoportban 12 tézist fogalmazott meg ezek közül kettőt, a 2.1 és 2.5 téziseket nem fogadom el, a 3.1 és 3.2 téziseket pedig egybe olvasztva, egy tézisként fogadom el. A további 8 tézist viszont fontos tudományos eredménynek tekintem és elfogadom.

Összefoglalva, az értekezés alapján a doktori munka tudományos eredményeit elegendőnek tartom az MTA doktori cím megszerzéséhez, és javaslom a nyilvános védelem kitűzését.

Budapest, 2024. február 14.

Kacsuk Péter (az MTA doktora)