

Bírálati vélemény

Toka László:

„Resource provisioning in cloud systems”

(Erőforrás-kezelés felhőrendszerekben)

című doktori értekezéséről

Az elmúlt két évtizedben a felhőrendszerek –azon belül is a hatékony erőforráskezelés– már számos tudományos kutatás tárgyát képezte világszerte. Ennek ellenére újabb és újabb kihívásokra kell választ találni tekintettel az olyan új technológiák és megközelítések folyamatos megjelenésére, mint például a perem- és ködszámítások (edge/fog computing) vagy a mikroszolgáltatás és 5G architektúrák. A jelölt által kitűzött célok is ezen témakörökbe tartoznak, időszerűek, sok esetben pedig már a jövőbeli, hosszabb távú, nagyszabású fejlesztéseket is megalapozhatják (lásd 1. téziscsoport eredményei).

A doktori értekezés 96 oldalon keresztül ismerteti az elért eredményeket, melyből 4 oldal bevezető, majd a 3 téziscsoport részletes kifejtése történik 86 oldal terjedelemben, végül a jelölt 6 oldalon összefoglalja a tudományos eredményeit. A bibliográfia 146 referenciát sorol fel. A weboldalakra hivatkozásoknál formai szempontból hiányolom az utolsó látogatás dátumának, illetve több konferenciakiadványnál az oldalszámok feltüntetését.

A doktori értekezés alaposan szerkesztett, gördülékenyen olvasható és összesen 30, a megfogalmazottak megértését jól segítő, minőségi ábrával illusztrált.

Az első két téziscsoportban 5-5, a harmadik téziscsoportban pedig további 2 tézist mond ki a jelölt.

A téziscsoportok a megcélzott tématerületet rendkívül széleskörűen átölelik: kezdve a felhők közötti, majd az adott felhőn belüli, végül pedig a felhasználók felhő-hozzáféréseivel kapcsolatos erőforráskezelési módszereire összpontosítva. Ez az igen széles, horizontális megközelítés tekinthető egyrészt erősségének is a doktori értekezésnek, de magában hordozza annak a veszélyét is, hogy az elért eredmények önmagukban némely esetben kevésbé mélyek vagy megalapozottak.

A választott kutatási módszereket és eszközöket tekintve egyértelműen kifejeződik az értekezésben a modellalkotásra és –általánosságban– a formális matematikai megközelítésre törekvés. Továbbá elmondható, hogy az így megválaszolt kérdések előkészítik a későbbi gyakorlati hasznosítási lehetőségeket, akár már csak ha az egyes kidolgozott algoritmusok polinomiális komplexitásának bizonyítására tekintünk (lásd 1. és 2. téziscsoport).

Az alkalmazott kutatási és matematikai módszerek (hasonlóan a lefedett témakörökhöz) szintén nagyon széles skálán mozognak, melyeket különböző mélységig aknázott ki a jelölt –többek között– játékelmélet (Stackelberg-játékok), gráfelmélet, gépi tanulás, valószínűségszámítás, sorbánállási modellezés, differenciálegyenlet-rendszerek, dinamikus programozás, Karp-redukció segítségével.

Az **1. téziscsoport** a felhők közötti üzleti kapcsolatok modellezésével foglalkozik egy lehetséges jövőbeli 5G ökoszisztémában, különös tekintettel az árazásra.

A jelölt az 1.1 tézisében jól definiált megszorítások mentén kialakított gráf topológiában összekapcsolt szolgáltatók és végfelhasználók esetén játékelméleti alapon adott felső limitet a közvetítői árakra a topológia egyensúlyi állapotához.

Kérdés (1): Milyen hatásai lehetnek a probléma megoldására, ha egy játékos (tipikusan egy közvetítő) dinamikus új kapcsolatokat is létre tud hozni?

Az 1.2 tézisben a jelölt 3 szintűre szigorította a lehetséges gráftopológiát, és így meghatározta mind a legfelső, mind pedig az egy szinttel alatta lévő szolgáltatók közvetítői árait Stackelberg-játékként értelmezve a problémát. Az 1.3 tézisben a jelölt bemutatja, hogy az ügyfelek NP nehéz problémával néznek szembe, ha tetszőleges a szolgáltatók és felhasználók kapcsolatát leíró gráf. Figyelembe véve, hogy NP nehéz problémáról van szó, a jelölt az 1.4 tézisben ismételt játékelmélet alkalmazásával sikeresen kifejezett (többek között) egyensúlyi árakat olyan egyszerűbb topológiákra, mint a párhuzamos utak soros hálózatokkal vagy fenyőfa topológia.

Kérdés (2): Hogy látja a széles körben elterjedt CloudSim alapú vagy más hasonló felhő/perem/ködszámítási rendszerek szimulációjára alkalmas eszköz, valamint a disszertáció különböző téziscsoportjaiban (pl. játékelméleti vagy más módszerekkel) elért eredmények egymás közötti viszonyát, különös tekintettel azok alkalmazhatóságainak határaitra, esetleges átfedő vagy kiegészítő jellegére?

Végül az 1.5 tézisben a jelölt általánosította az egyensúlyi árakra vonatkozó eredményeit a végtelen kapacitással rendelkező felhő- és hálózati szolgáltatók esetére.

A **2. téziscsoportban** már a nagyléptékű peremfelhő rendszerekre helyeződik át a fókusz: a magas megbízhatóság és egyidejűleg az alacsony késleltetés biztosítását célozza meg a jelölt a gazdaságossági szempontok figyelembevételével. Az online és offline erőforrásütemezők ún. helyőrzőket alkalmaznak, továbbá kidolgozásra került egy csomópont-szegmentáló eljárás – különös figyelmet fordítva a determinisztikus végrehajtás problémakörére. A 2.1 tézisben a jelölt bizonyította, hogy mindhárom bemutatott algoritmus polinomiális komplexitású. A 2.2 tézisben pedig bizonyításra kerül, hogy a javasolt online ütemezési módszer a helyőrzők együttes elhelyezésére egy 3-approximációs algoritmus.

Kérdés (3): A bemutatott erőforrásütemező algoritmusokat egyszerre egy csomópont meghibásodására készítette fel a jelölt. Felmerül a kérdés, hogy kiterjeszthető-e a megközelítés azokra az esetekre is, amikor egyszerre akár több csomópont esik ki – szem előtt tartva azt, hogy a jelölt hangsúlyozza, hogy a téziscsoport a nagyléptékű peremfelhők kutatási területét célozza meg.

A téziscsoport a továbbiakban a gépi tanulás alapú automatikus skálázás lehetőségeit vizsgálja felhő alkalmazásokra, és a 2.3 tézisben a gyakorlati életben is elterjedt Kubernetes orkesztrációs keretrendszer automatikus skálázó egységét (HPA) írja le formális módszerekkel.

Kérdés (4): Az autóskálázás támogatására a disszertációban bevezetésre kerül az alkalmazás profilozás (lásd 3.4 és 3.5 ábra), de a Kubernetes automatikus skálázó analitikus modelljébe már ezen paraméterek konstans értékkel kerülnek be (55. oldal). Mennyire befolyásolhatja ez az elért eredmények gyakorlati alkalmazhatóságát a többi megszorítással egyetemben, mint például, hogy a leskálázási (downscale) stabilizáció egyenlő a skálázási intervallummal?

A jelölt ezután a 2.4 tézisben felhőben telepített alkalmazásokhoz mutat be egy saját gyűjtésű adatok felhasználásával kidolgozott gépi tanulás alapú ensemble modellt négy elterjedt modell előnyeinek ötvözésével (AR, LSTM, HTM, RL), majd pedig az arra épülő új, prediktív automatikus skálázási módszert (HPA+).

Kérdés (5): Tekintettel a megfelelő minőségű, mennyiségű és kellően reprezentatív tanítási adathalmaz kiemelt fontosságára a kutatás során, a jelölt véleménye szerint befolyásolhatja-e az eredmények szélesebb körű értelmezését a munkája során használt idősoros tanítási adatok, melyek 3+1 munkanapot felölelve, egy alkalmazásra vonatkozóan és egy egyetemi hálózaton keresztül kerültek begyűjtésre. Ezzel összefüggésben mennyire tekinthetőek ezek az adatok a tézisben is említett benchmark eredmény (vagy más, a szakirodalomban bemutatott módszerekkel lehetséges összehasonlítás) megbízható alapjának?

Végül a téziscsoport utolsó, 2.5 tézisében a jelölt ismertet egy módszert a szükséges erőforrás-többlet minimalizálásra a nagyléptékű alkalmazás skálázási csoportokba rendezéséhez.

Kérdés (6): Az értekezés alapvetően diszkrét idejű leírással és modellekkel foglalkozik, így felvetődik a kérdés, hogy feltétlenül szükséges-e áttérni folytonos idejű analitikai módszerekre, parciális differenciálegyenletek bevezetésére, vagy más lehetőségek is alkalmazhatóak itt?

A **3. téziscsoportban** ismét tolódik a fókusz, a jelölt immár egy koordinációmentes és egy koordinált erőforrás-allokációs sémát ismertetett a hálózati sávszélesség felosztására a felhőalkalmazások ugyanazon hozzáférési hálózaton lévő kliensei között. A 3.1 tézisben numerikus elemzéssel bemutatásra került, hogy meghatározott körülmények között egy adott, bemutatott stratégiák egyike sikeresen alkalmazható.

Kérdés (7): A Markov döntési folyamat (MDP) bevezetése kapcsán a vizsgálat során felmerült memóriaigény problémák milyen alkalmazás-paraméterek (pl. kliensek száma) esetén válnak gyakorlati szempontból kezelhetetlenné, illetve hogy látja a kapott eredményei verifikálhatóságát (bizonyos keretek vagy megszorítások mellett) MDP modell segítségével?

A 3.2 tézisben a jelölt már egy koordinált erőforrás-allokációs sémát mutatott be és bizonyította, hogy milyen komplexitású a kapcsolódó dinamikus programozást alkalmazó algoritmus, amit az optimumszámításhoz vezetett be.

Minden téziscsoport esetén elmondható, hogy megtörtént a szakirodalom viszonylag részletes feldolgozása, de az olvasónak hiányérzete lehet az olyan hazai kutatási eredményekkel történő összevetésre vonatkozóan (akár kapcsolódásuk, akár kiegészítő jellegük folytán), melyek a felhő-, perem- és ködszámítási rendszerek szimulációs és orkesztrációs módszerei területén születtek, illetve szintén a felhő-erőforráskezelés területét célozzák meg általánosságban. Továbbá R. Buyya kutatócsapatának eredményeivel (University of Melbourne) történő átfogóbb összehasonlítás növelte volna a benyújtott értekezés színvonalát.

A tézisek alátámasztására szolgáló tudományos közlemények részben magas IF értékkel rendelkező folyóiratokban (pl. IEEE Transactions on Network and Service Management), részben pedig rangos nemzetközi IEEE vagy ACM konferencia kiadványokban jelentek meg 2017 és 2022 között (jóval a PhD megszerzését követően), ahol a jelölt első- vagy utolsó szerzőként került feltüntetésre. Az értekezés erőssége, hogy a tudományos publikációk mellett a 2. téziscsoporthoz 1 db USA-ban benyújtott (függőben lévő) szabadalmat, a 3. téziscsoporthoz pedig 2 db már az USA-ban bejegyzett szabadalmat tüntet fel a jelölt a művében.

Az eredmények hasznosíthatósága a bemutatott projektek, szabadalmak és egyéb munkák alapján példaértékűnek mondható.

A fentiek alapján a jelölt –megítélésem szerint– saját eredményeit mutatta be az értekezésben, hiteles adatokat tartalmaz.

A bemutatott téziseket új tudományos eredményként elfogadom, de az 1.3 és 2.4 tézisek kapcsán megjegyezném, hogy az 1.3-as tézis ugyan fontos állítása a téziscsoportnak, bár bizonyítása kézenfekvőbb, mint a többi esetén, a 2.4-es tézis esetén pedig a tanító adathalmaz mérete, minősége és reprezentativitása ronthat az értekezésben bemutatott tudományos eredmények egyenszilárdságán és jelentőségén.

Összefoglalva a bírálatot: a jelölt alapos, széleskörű és értékes kutatómunkát ismertetett értekezésében –néhány gyengébb, de kevésbé kritikus ponttól eltekintve– új tudományos eredményeket felmutatva egy rendkívül kurrens informatikai területen, ezért az értekezés nyilvános vitára bocsátását támogatom.

Budapest, 2024. január 25.

.....

Dr. habil. Lovas Róbert
tudományos főmunkatárs, egyetemi docens