



innovation informatics infocommunications

ELTE
PPKE
BME

**Innovatív informatikai és
infokommunikációs megoldásokat
megalapozó tematikus kutatási
együttműködések 3.**

EFOP-3.6.2-16-2017-00013.

SZÉCHENYI 



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

Innovatív informatikai és infokommunikációs
megoldásokat megalapozó tematikus kutatási együttműködések 3.

Budapesti Műszaki és Gazdaságtudományi Egyetem, 2019

Szerkesztők:

Dr. Jereb László
Dr. Dudás Ákos
Gerhátné Dr. Udvarý Eszter
Dr. Szűcs Gábor

ISBN 978-963-421-793-0

**Innovatív informatikai és infokommunikációs
megoldásokat megalapozó
tematikus kutatási együttműködések
3.**

EFOP-3.6.2-16-2017-00013.

Tartalom

Előszó	4
Szoftverfejlesztés és informatikai biztonság (A pillér)	7
Dinamikus metamodellezés	10
Dinamikus modellezés	11
Virtuális gép dinamikus metamodellezéshez	12
Gráfadatbázisok teljesítményelemzése	13
Programelemzési módszerek alkalmazása	
kártékony programok azonosítására és vizsgálatára	14
Trigger-alapú viselkedés detektálása bináris futtatható kódokban	15
Okos mérők biztonsági tesztelése	16
Malware minták hasonlóságának vizsgálata	17
Automatizált protokoll visszafejtés a hálózati forgalom statikus elemzésével	18
Konvergens IoT megoldások – Egységes szoftverfejlesztési módszertan	19
Az Ipar 4.0 és Ipari IoT platformok megismerése	20
IoT a gyógyszergyártásban	21
Hálózati forgalom vizsgálatára épülő adatvédelem az okosotthonban	22
Anonimizálás folyamatosan érkező adatokban	23
Anonimizálás a Differential Privacy modell és a k-anonimitás kombinálásával	
a GDPR-nak való megfelelés érdekében	24
Big Data SQL végrehajtó automatikus választása machine learning alapokon	25
Mesterséges intelligencia alkalmazása GDPR-nak megfelelő adatfeldolgozás során	26
Szolgáltatásbiztonság garantálása kritikus CPS rendszerekben	27
Modellalkotás mérési adatokból	28
Dinamikus rekonfiguráció edge alapú kritikus CPS rendszerekben	29
Kritikus CPS rendszerek formális módszerekkel támogatott analízise	30
Demonstrátorok szolgáltatásbiztos CPS rendszerek bemutatására	31
Infokommunikációs hálózatok és kiberfizikai rendszerek (B pillér)	33
Valós idejű CPS rendszerek	36
IEEE 1588 mesteróra fejlesztése kísérleti CPS rendszerekhez	37
Környezeti monitorozó rendszer levegőtminőség vizsgálatára	38
Adatvesztési modellek vizsgálata idő- és frekvenciatartományban	39
Felhő alapú hálózati megoldások	40
Serverless megoldások teljesítményvizsgálata	41
Function as a Service alapú távvezérlés	42
Nyílt forráskódú Function as a Service platformok optimalizálása	43

Számítási felhő és minősége	44
Ütemezési algoritmusok vizsgálata	45
Szoftverrádió-alapú IoT fejlesztőrendszer	46
Masszív MIMO feldolgozás IoT alkalmazásokban	47
Időkritikus 5G architektúrák	48
Járművek közti kommunikáció látható fény segítségével	49
Valós idejű bitgenerálás kvantum-véletlenszámgenerátorral	50
Neurális hálózatok alkalmazása fading jelenségek automatikus detekciójában	51
A digitális infrastruktúra szolgáltatási képességeinek fejlesztése és automatikus üzemeltetésének megalapozása	52
Erőforrás felhasználásának optimalizációja aktor-alapú rendszerekben	53
Adatközponti hálózati topológiák hatása virtuális hálózati funkciók szolgáltatásláncainak megvalósítására	54
Véletlen hibák okozta hálózati vészhelyzeti incidensek kezelése szolgáltatói IP hálózatokban	54
Járműkommunikációs hálózatok tervezési, elemzési, szimulációs modelljei és módszerei	55

Intelligens adatelemzés (C pillér)	57
---	-----------

Gyengén koherens sokdimenziós idősorok modellezése mély tanulással	60
Hierarchikus Temporális Memória alapú idősor modellek számításigényének csökkentése	61
Ajánlati könyvek modellezése mély tanulással	62
Intelligens adatelemzés kiberfizikai (CPS) rendszerekben	63
Intelligens gépkocsik városi forgalomban – kommunikáción alapuló kooperáció vizsgálata	64
Neurális hálók komplexitásának automatizált redukciója	65
Képszegmentálás városi környezetre jellemző objektumok detektálására önműködő járművekhez	66
Nagy adathalmazok (Big Data) komplex struktúráinak előrejelzése az adattudomány módszereivel	67
Szövegdokumentumok összefoglalóinak automatikus előállítás	68
Aktív tanulás és spektrális klaszterezés nagy adathalmazoknál	69
Implicit visszacsatoláson alapuló kollaboratív szűrős ajánlórendszer mély neurális hálóval	70
Döntéstámogató módszerek dinamikus orvosi képalkotó berendezésekhez	71
Neurális hálózattal támogatott PET rekonstrukció szinogram adatokból	72
PET rekonstrukció javítása szűrővel és mozgáskompenzációval	73

Előszó

Jelen összeállítás a **3IN**-nek rövidített, „Innovatív Informatikai és Infokommunikációs megoldásokat megalapozó tematikus kutatási együttműködések” című, EFOP-3.6.2-16-013 azonosítójú projekt 3. mérőldkövének végéig elért BME eredményeket foglalja össze. A kutatásokat a **Szoftverfejlesztés és informatikai biztonság** (A/ pillér), az **Infokommunikációs hálózatok és kiberfizikai rendszerek** (B/ pillér), valamint az **Intelligens adatelemzés** (C/ pillér) témakörökben folytattuk, amelyeket rendre az ELTE (A), a BME (B) és a PPKE (C) munkatársai koordinálnak.

Az elmúlt több mint kettő évben 6 tanszéken témavezetők és mentorok támogatásával 120 PhD, MSc és BSc hallgató kapott ösztöndíjat, akik a 15 BME témában a három pillér között közel egyenletesen (5+6+4) kutatnak. A projekt keretében elért hallgatói eredmények 60 angol nyelvű folyóiratcikkből és konferenciakiadványban jelentették meg, valamint 100 több külföldön és hazánkban rendezett eseményen (konferencián, workshopon stb.) adták elő angolul. A hozzájárulások listáját 130 további dokumentum, TDK dolgozat és előadás, diplomaterv, szakdolgozat, részletes kutatási jelentés, valamint e kiadványsorban publikált rövid összefoglaló színesíti.

A projektben a BME egyrészt az ELTE és BME több éve sikeres EIT Digital tevékenységének erősítésére vállalkozott, másrészt hozzá kívánunk járulni a balatonfüredi térség, általánosabban Veszprém megye, a Közép-dunántúli régió gazdasági fejlődéséhez is. E cél elérése érdekében hallgatóink minden mérőldkö végén Balatonfüreden mutatják be az eredményeket, így az 1. mérőldkö zárórendezvényét 2018. június 18-án, a 2. mérőldkö workshopját pedig 2019. január 30-31-én rendeztük. Jelen kiadvány a 2019. november 22-23-i eseményre jelenik meg és – döntően az utolsó évben elért eredményekkel – azt mutatja be, hová jutottak el az egyes témák művelői.

A dokumentum követi az 1. és 2. mérőldkö végén megjelent kiadvány felépítését. Sorra veszi az egyes pilléreket, magyar és angol nyelven röviden összefoglalja azok fő célkitűzéseit. Pillérenként az egyes témák egyoldalas magyar/angol összegzését követően magyar nyelven, 1-1 oldalon az adott téma 1-4 résztémájában végzett hallgatói kutatások célját, módszereit és eredményeit ismerteti, általában esetben megjelent publikációkra is hivatkozva. Az egyes pillérek és témák esetében megadjuk a pillér koordinátorok és témavezetők elérhetőségét, az egyes résztémák esetében pedig a kutatási tevékenységet végző hallgatók, valamint a munkájukat segítő mentorok és/vagy témavezetők nevét is.

Ezen kiadvánnyal is az a célunk, hogy színesítsük a projektben végzett munkáról és BME eredményekről bemutatott eddigi képet, miközben tovább kívánjuk erősíteni a széles értelemben vett balatonfüredi térség vállalataival a tartalmas munkakapcsolatokat is.

Foreword

This compilation summarizes the BME results achieved by the end of Milestone 3 of the project EFOP-3.6.2-16-013, “Thematic Research Collaborations for Innovative Informatics and Information Solutions” abbreviated to 3IN. The research was carried out in the areas of Software Development and Information Security (A / Pillar), Infocommunication Networks and Cyber-physical Systems (B / Pillar) and Intelligent Data Analysis (C / Pillar), that were coordinated by ELTE (A), BME (B) and PPKE (C), respectively.

Over the past two years, 120 Ph.D., MSc, and BSc students have been awarded scholarships researching 15 BME topics almost evenly (5 + 6 + 4) distributed among the three pillars with support of supervisors and mentors in 6 departments. The achievements of the students have appeared in more than 60 English papers in journals and proceedings, in more than 100 presentations delivered at conferences and workshops organized abroad or in Hungary. The list of other contributions contains about 130 TDK papers and talks, MSc and BSc theses, detailed research reports, and short summaries published in the previous edition of this series.

On the one hand, BME has reinforced the successful EIT Digital activities of ELTE and BME for several years with the help of the current project. On the other hand, we intend to contribute to the economic development of the Balatonfüred region, more generally Veszprém county and the Central Transdanubian region. The students present their results at the end of each milestone in Balatonfüred, such as at the closing event of the 1st milestone on June 18, 2018, and at the workshop of the 2nd milestone on January 30-31, 2019. This document will be published on the 3rd workshop on November 22-23, 2019, and will present the status of the research carried out by each thematic team, focusing on the results achieved in the last year.

The compilation follows the structure of the previous ones published at the end of Milestones 1 and 2. It briefly summarizes the primary objectives of the pillars in Hungarian and English. Following the one-page summary of each research topic in Hungarian and English, one column per page describes the aims, methods, and results of the individual student research subtopics in Hungarian, referring in many cases to publications for details. We provide contact details for the pillar coordinators and supervisors and each sub-topic. The names of the students are given with the supervisors and mentors assisting them in their research.

The objective of the publication is to enhance the further impact of results achieved by BME researchers so far while it also strengthens the relationship with companies located in the broader region of Balatonfüred.



Dr. László Jereb

Egyesült Innovációs és Tudásközpont
Federated Innovation and Knowledge Centre
jereb.laszlo@eit.bme.hu

Szoftverfejlesztés és informatikai biztonság

(A pillér)

Szoftverfejlesztés és informatikai biztonság

(A pillér)

A pillér keretében a BME-n öt témakörben folyik kutatás hallgatók bevonásával és munkájuk folyamatos mentorálásával. A pillérhez kapcsolódó kutatás a szoftverfejlesztés és az ehhez kapcsolódó biztonság területén változatos témaköröket ölel fel. A *Dinamikus metamodellezés* témában (TA.21) egy új modellezési módszertan kidolgozása a cél, ami lehetővé teszi, hogy a kezdeti, magas absztrakciós szinten meglévő specifikációtól kezdődően egészen a végleges implementáció elkészítéséig egy egységes, mindvégig validált, de mégis rugalmas környezetben hozzassunk létre rendszereket. Jelen mérőföldkő során az eredmények tudományos publikálása mellett a praktikus felhasználhatóság került fókuszba egy vizuális szerkesztőfelülettel és gyakorlati megvalósítási kérdésekkel. A *Programelemzési módszerek alkalmazása kártékony programok azonosítására és vizsgálatára* téma (TA.22) keretein belül a cél megvizsgálni, hogy a szoftver termékekben milyen kártékony, nem dokumentált funkciók találhatók meg, amelyek a rendszer működését károsan befolyásolhatják. Módszereket fejlesztettünk kártékony viselkedést triggerelő inputok automatizált azonosítására, sérülékenységek azonosítására beágyazott eszközökön és azok protokolljaiban, és ismeretlen protokollok automatizált visszafejtésére. Foglalkoztunk továbbá malware minták osztályozásával bináris hasonlóság alapján. A *Konvergens IoT megoldások – Egységes szoftverfejlesztési módszertan* (TA.23) téma olyan egységes szoftverfejlesztési módszertant biztosító elemek és folyamatok azonosításával, kidolgozásával és alkalmazásával foglalkozik, amelyek a konvergens szoftver megoldások irányába mutatnak. Ennek részeként megvizsgálásra kerültek népszerű IoT platformok, és konkrét IoT-alapú gyártási folyamat ellenőrző eszköz implementációk is készültek. Az *Anonimizálás folyamatosan érkező adatokban* (TA.24) témában az alapvető kérdés, hogy hogyan lehet a kezdetben nem teljes mértékben rendelkezésre álló, folyamatosan érkező adatok esetében az anonimizálási feladatokat ellátni. A kutatásunk célja egy korábban kifejlesztett módszer továbbgondolása annak érdekében, hogy a módszer a GDPR előírásainak is megfelelő módon alkalmazható legyen. Emellett fő célunk a módszer oly módú kibővítése volt, amely a lehető legkevesebb érzékeny adatot bízta az adatkezelőre, helyette az adatgyűjtés helyén végzi el az anonimizálást. A *Szolgáltatásbiztonság garantálása CPS rendszerek* (TA.25) téma célja az esetenként biztonságkritikus CPS funkciók biztonságosságát garantáló tervezési és futási idejű validációs és verifikációs, valamint hibatűrést biztosító megoldások kidolgozása. Célunk a szabványosodó, Edge alapú CPS architektúrák feletti hibátűrő minták modell alapú kidolgozása, és benchmark mérésekből vagy működési naplókából teljesítőképességi és szolgáltatásbiztonsági modellek és analízis megoldások kidolgozása, esetenként a gépi tanulás eszközeit is integrálva.

Software development and IT security

(Pillar A)

This pillar is divided into five topics in which the research is applied by involving students and their mentors. The research covers various topics in the field of software development and its security aspects. In the topic of *Dynamic metamodeling* (TA.21), the goal is to create a new modeling method, which enables the creation of systems in a flexible yet validated environment all through the development lifecycle starting from the high-level specification and ending in the final implementation. In this phase, the focus was on the publication of the scientific results, as well as moving forward in the direction of practical applicability, including the development of a visual editor. The topic *Detection and analysis of malicious software using program analysis methods* (TA.22) the aim is to investigate how program analysis methods can be used to detect malicious, non-documented components and functions in computer software. We developed an automated method that makes it possible to identify inputs that trigger malicious behavior in a malware, we developed methods for detecting vulnerabilities in embedded firmware and protocols, and we continued our work on automated protocol reverse engineering. In addition, we studied the problem of malware classification based on binary similarity of malware samples. The topic *Convergent IoT Solutions – Unified Software Development Methodology* (TA.23) focuses on analyzing, elaborating and applying a unified software development methodology with the goal to step forward to convergent software solutions. Our work examined popular IoT platforms, and on a parallel track, implemented IoT-based production line quality assurance tools. The basic question in the topic of *Anonymization of stream data* (TA.24) is how we can apply anonymization techniques in streaming data not available at the beginning of processing. We focused on enhancing an existing anonymization method so that it is applicable for GDPR-compliant data processing. Furthermore, the goal of this enhancement is to reduce the amount of privacy-sensitive information data processors have to handle by moving the burden of anonymization to the data-gathering site. The goal of the topic *Assuring dependability in cyber-physical systems* (TA.25) is the elaboration of validation and verification methods both in design and run-time. The research focuses on fault-tolerant design patterns over edge-computing based CPS architectures. Performability and dependability model identification and validation, starting from benchmark experiments or operational logs supported by machine-learning, is the primary paradigm used.



Dr. Dudás Ákos

Automatizálási és Alkalmazott Informatikai Tanszék
Department of Automation and Applied Informatics
Dudas.Akos@aut.bme.hu

Dinamikus metamodellezés

A dinamikus metamodellezés téma ígéretes alapkutatásként indult, ami mára bizonyította, hogy az önleíró struktúrája, valamint az egyedülálló validált, rugalmas koncepció finomítási képessége segítségével ígéretes technológia lehet a gyakorlatban is az Ipar 4.0 világában. A kutatás harmadik fázisában egyidejűleg fókuszáltunk az elméleti háttér teljesebb kifejtésére, valamint a praktikus felhasználhatóság elérésére. Elméleti szempontból az eredmények a tudományos közéletben történő elhelyezése és publikálása kapott nagy hangsúlyt, míg gyakorlati szempontból több ágra vált szét a munka. Egyrésztől megszületett egy vizuális szerkesztőfelület prototípusa, ami a szerkesztést megkönnyíti és hasonló célból született egy webes szerkesztő alkalmazás is. A dinamikus felhasználás érdekében áttértünk a GraalVM-alapú virtuális gépre, ami több új lehetőséget is hozott. Végül a tárolás kapcsán elkezdtuk felmérni a létező gráfadatbázisok képességeit, hogy kiválaszthassuk a legalkalmasabb platformot az adatok tárolására.

Dynamic metamodeling

The dynamic metamodeling topic started as a promising new fundamental research direction, which has proved its usefulness in the age of Industry 4.0 based on its self-describing structure and unique, validated, but flexible stepwise refinement ability. During the third phase of the research, we have focused simultaneously on completing the theoretical background and the practical applicability. From the theoretical point of view, publications of the results gained the most attention, while from the practical viewpoint, the work has split into several branches. A prototype of the visual editor and a web-based editor have been created to simplify the editing of the models. In order to support the dynamic behavior, we have moved to a Graal-based virtual machine instead of the old, compiler-based architecture. Finally, we have started to discover the main properties of the most popular graph databases in order to find the right candidate to store the data.



Dr. Gergely Mezei

Automatizálási és Alkalmazott Informatikai Tanszék
Department of Automation and Applied Informatics
gmezei@aut.bme.hu

Dinamikus modellezés

Hallgatók: Bácsi Sándor, Palatinszky Dániel

Mentor: Dr. Mezei Gergely

A dinamikus metamodellezés témakörén belül fontosnak tartjuk az eredmények összehasonlítását, elhelyezését a hasonló kutatási eredmények körében. Ennek keretében belül a DMLA által támogatott funkciókat összevetettük a klasszikus objektumorientált paradigmával, összegyűjtöttük a hasonlóságokat, illetve a különbségeket [1]. A többszintű metamodellezés terén is alapos kutatómunkát végeztünk [2]. Az eredmények alapján sikerült megérteni és feltárni a DMLA által támogatott többszintű megközelítés valódi előnyeit a kutatási területen belül megjelenő más elképzelésekhez képest. A DMLA elemzéséhez kapcsolódóan újfent áttekintettük a koncepció alapkövének tekinthető Bootstrap felépítését és egy cikk [3] formájában publikáltuk a főbb tervezési döntéseket, a koncepciók kapcsolatát és az alapvető építőköveket.

A gyakorlati, felhasználói oldalt tekintve a DMLA-ban használt koncepciók tesztelésére és illusztrálására elkészült egy EMF és Sirius alapú vizualizációs modellező felület, amely egy magasabb absztrakciós szinten biztosítja a DMLA modellek elkészítését és azok refaktorálását [4].

Az implementációs vonulat mellett sikerült nemzetközi szinten is fontos publikációs eredményeket elérni, amelyeket az idei MODELS 2019 konferencia workshopjain mutattunk be. Kidolgoztunk egy algoritmust a korábban problémát jelentő öröklési kapcsolatok megjelölésére, megoldást küldtünk a MULTI Workshop 2019-es Process Challenge nevű kihívására, ami kifejezetten a többszintű metamodellezési módszerek képességeinek összehasonlítására született, valamint új megközelítésként publikáltunk egy elemzést a DMLA rugalmasságát tekintve a FlexMDE workshopon.

Referenciák

- [1] BÁCSI S., MEZEI G., THEISZ Z., SOMOGYI F., URBÁN D.: *Comparing multi-level modelling and the classic object-oriented method*. In: Conference on Software Technology and Cyber Security: STCS 2019 (2019).
- [2] F. A. SOMOGYI, S. BÁCSI, G. MEZEI, Z. THEISZ, D. URBÁN: *Challenges in the field of multi-level modeling*. Automation and Applied Computer Science Workshop 2019.
- [3] G. MEZEI, F. A. SOMOGYI, Z. THEISZ, S. BÁCSI, D. PALATINSZKY, D. URBÁN: *A bootstrap for self-describing, self validating multi-layer metamodeling*. Automation and Applied Computer Science Workshop 2019.
- [4] Sándor BÁCSI: *Dynamic Metamodeling*. Diploma Thesis 2019.

Virtuális gép dinamikus metamodellezéshez

Hallgató: Palatinszky Dániel

Mentor: Dr. Mezei Gergely

A kutatócsoportunk által kitűzött célok egyike, hogy a Dynamic Multi-Layer Algebra (DMLA) többretegű metamodellezési keretrendszer rugalmasságát, dinamikusságát és teljesítményét növeljük. Ennek keretein belül a korábbi fordítás-alapú megközelítését kiváltottuk egy sokkalta dinamikusabb interpretálás alapú megvalósítással [1]. Az interpretálás lehetővé teszi a DMLA saját programozási nyelvével, a DMLAScripttel készített modellek valósidejű módosítását és validálását, ezáltal lehetőséget teremtve vizuális szerkesztőeszközök készítésére és a keretrendszer felhasználóbarát kezelésére. Az interpreter megvalósításához egy új generációs Java virtuális gépet és fordítót, a GraalVM-et használjuk. A GraalVM kitüntetett célja egy olyan nyelvfüggetlen futtatókörnyezet biztosítása, amely natívhoz közeli teljesítményt képes nyújtani.

Az új megközelítés első mérföldköve a fordítás alapú megközelítés funkcionálisának lefedése volt, melyet mostanra sikeresen elértünk. Ez egyrészt azt jelenti, hogy a modellek virtuális gépben való reprezentációja bizonyítottan megegyezik a fordítás alapú ábrázolással. Másrészt az új implementáció által adott validációs végeredmények minden egyes meglévő domain esetén megegyeznek a korábbi megvalósítás eredményeivel. Összességében tehát a jelenlegi implementáció egy biztos alapot nyújt a további fejlesztésekhez. Mindezen felül a teljesítményt is sikerült javítani, mely egyrészt köszönhető a GraalVM és Truffle által nyújtott optimalizációs lehetőségeknek, másrészt a fordítás alapú megvalósítás bizonyos feleslegessé vált folyamatainak elhagyásának. Eközben a projekt másik ágán folyó webes szerkesztőfelület fejlesztése során már át is álltunk a GraalVM-alapú virtuális motor használatára.

A kutatócsoportunk végső célja a DMLA keretrendszer iparban való felhasználhatóságának biztosítása, melynek első lépéseit sikeresen megtettük a dinamizmus kezdetleges megvalósításával és a keretrendszer szélesebb körű bemutatásával a multi-level metamodellezési világ felé [2].

Referenciák

- [1] Dániel PALATINSZKY, Dániel URBÁN, Dmitry DUNAIEV: *Introduction to multi-layer metamodeling with Graal and Truffle*. AACS'2019.
- [2] Zoltán THEISZ, Sándor BACSI, Gergely MEZEI, Ferenc A. SOMOGYI, Dániel PALATINSZKY: *By multi-layer to multi-level modeling*. MULTI 2019.

Gráfadatbázisok teljesítményelemzése

Hallgató: Kovács Tibor

Mentorok: Dr. Mezei Gergely, Simon Gábor

A Dynamic Multi-Layer Algebra (DMLA) megközelítés a lépésenkénti finomításra épülve igen összetett modellt eredményezhet egy ipari rendszer esetén, ahol a klasszikus adatbázisok tárolási formája nem megfelelő, hiszen a kapcsolatok mélysége igen jelentős lehet. A problémakör kezelésére intuitív választásnak tűnik a gráfadatbázisok használata.

Munkánk során megterveztünk és implementáltunk egy összetett, több faktort is figyelembe vevő benchmark keretrendszert, mely segíthet kiválasztani a célnak legmegfelelőbb eszközt. Ennek első lépéseként kiválasztottuk a manapság népszerű gráfadatbázisok egy halmazát, majd megvizsgáltuk az egyes gráf-alapú metaadat modellezési lehetőségeket és ezek implementálhatóságát az egyes rendszereken. Ezt követően kiértékelünk néhány adatbázis teljesítményét [1] mesterséges lekérdezések segítségével. Ennek részeként kialakítottunk egy a későbbiekben is felhasználható felhős mérési infrastruktúrát. Az eredmények kiértékelése során megállapítottuk, hogy adott típusú feladatokra melyik implementáció-metaadat modell páros az optimális választás.

A mesterséges lekérdezéseket követően különböző megközelítéseket felhasználva szintetizáltunk egy valós, életszerűnek tekinthető lekérdezés halmazt, majd ezt felhasználva elkezdtük a kiválasztott rendszerek teljesítményének több aspektusra kiterjedő részletes mérését. Ennek a mérésnek az első eredményeit a [2] publikáció tartalmazta, melyben a felhő-alapú és a dedikált, hagyományos futtatókörnyezetben futó gráfadatbázis-kezelőket hasonlítottuk össze az Azure Cosmos DB és a Neo4j eredményei alapján. A publikált eredmények csak a lekérdezés válaszütemét tartalmazták, azóta több más fontos paraméter is mérésre került, például a tárhely igény, vagy a skálázódási faktor, illetve további rendszerek is bevonásra kerültek, mint például a Blazegraph vagy a JanusGraph.

Referenciák

- [1] Tibor KOVÁCS, Gábor SIMON, Gergely MEZEI: *Benchmarking Graph Database Backends—What Works Well with Wikidata?* Acta Cybernetica 2019: 43-60.
- [2] Tibor KOVÁCS, Gábor SIMON: *Storing Large-scale Knowledge Bases in Cloud-native Graph Databases: Are They Up for the Task?* AACS 2019: 129-134.

Programelemzési módszerek alkalmazása kártékony programok azonosítására és vizsgálatára

A számítógépes rendszerek működése – a felhő alapú szolgáltatásoktól, a mobil alkalmazásokon keresztül, a kiberfizikai beágyazott rendszerekig – szoftverekre épül. Ezért a káros funkciót megvalósító szoftver-modulok jelentős negatív hatással vannak rendszereink biztonságára. Ebben a projektben célunk annak vizsgálata, hogy programelemzési módszerek segítségével felderíthetők-e, és ha igen, akkor hogyan és milyen hatékonysággal, egy szoftverben elhelyezett, nem dokumentált, kártékony funkciók. Kifejlesztettünk egy módszertant beágyazott eszközök biztonsági tesztelésére, amit az elmúlt évben egy okosmérő berendezés és kapcsolódó szoftverkörnyezet elemzésére használtunk. Kifejlesztettünk továbbá egy szimbolikus programvégrehajtásra épülő módszert, amivel automatizáltan lehet azonosítani egy rosszindulatú szoftverben a kártékony funkciót triggerelő inputokat. Folytattuk az ismeretlen protokollok automatizált visszafejtését, ami segíti például egy kártevő által folytatott távoli kommunikáció megértését. S végül, vizsgáltuk malware minták osztályozását bináris hasonlóság alapján, ahol a hasonlóság mérésére a TLSH nevű fuzzy hash algoritmust használtuk.

Detection and analysis of malicious software using program analysis methods

The operation of contemporary computer systems – including cloud-based systems, mobile applications, and embedded cyber-physical systems – is largely based on software. Hence, malicious software modules can have a strong negative impact on the security of these systems. In this project, our aim is to investigate how program analysis methods can be used to detect such malicious, non-documented components and functions in computer software. We developed a penetration testing methodology for embedded systems, which uses static program analysis techniques and protocol reverse engineering techniques to identify vulnerabilities. We also developed an automated method based on symbolic program execution for identifying inputs that trigger malicious behavior in malware. We continued our work in the field of automated protocol reverse engineering which can be very helpful in understanding the communication of malware samples with their command and control centers. Finally, we studied the problem of malware classification based on binary similarity of malware samples, where similarity was measured by the TLSH fuzzy hash algorithm.



Dr. Buttyán Levente

Hálózati Rendszerek és Szolgáltatások Tanszék
Department of Networked Systems and Services
buttyan@hit.bme.hu

Trigger-alapú viselkedés detektálása bináris futtatható kódokban

Hallgató: Futóné Papp Dorottya

Mentor: Dr. Buttyán Levente

Trigger-alapú viselkedésnek nevezzük azt a jelenséget, amikor egy szoftver egy nem dokumentált, sokszor rosszindulatú viselkedést csak bizonyos körülmények között hajt végre. Rosszindulatú támadók számára nagyon előnyös ilyen viselkedést implementálni a rosszindulatú szoftverekbe, mivel nehezen detektálható, rejtőzködő rosszindulatú szoftvereket készíthetnek e technika felhasználásával. Például a rendszeridő kiolvasásával időzíthetik, illetve a hálózaton keresztül küldött parancsokkal kontrollálhatják a támadást. Lehetőségük van a rendszeren éppen futó folyamatok lekérdezésével ellenőrizni, hogy elemző környezetben futtatják-e a rosszindulatú szoftvert, és ha igen, akkor nem indítják meg a támadást, ezáltal elkerülve a detekciót.

Elemzési szempontból nehéz kihívást jelent egy olyan rosszindulatú szoftver, amely trigger-alapú viselkedést implementál. Csak akkor tudjuk elemezni a rosszindulatú viselkedést, ha ismerjük a triggereléséhez szükséges környezet beállítását. Ehhez azonban programelemzési és -visszafejtési technikák alkalmazására van szükség, amiket manapság általában manuálisan végeznek el.

Jelen kutatásban a triggereléshez szükséges környezet automatizált összeállításával foglalkoztunk. Kidolgoztunk egy új módszert [1], amely képes elérni egy, az elemző által kijelölt érdekes utasítást a bináris kódban és kinyerni a futtatásához szükséges környezet beállításait. A módszer alkalmazhatóságát egy valós életből vett rosszindulatú kódon, a Kaiten családból származó mintán teszteltük. A minta denial of service (DoS) támadásokat indít, amennyiben a megfelelő távoli parancsot kapja a támadótól. A módszerünk képes volt elfogadható időn belül elérni egy általunk kiválasztott DoS támadás implementációját és rekonstruálni a támadás triggereléséhez szükséges parancsot. A módszerünk az első lépés afelé, hogy képesek legyünk kontrollált környezeteket automatizált módon összeállítani úgy, hogy abban elemezhetővé váljanak korábban rejtett, potenciálisan kártékony viselkedések is.

Referencia

- [1] PAPP D., TARRACH T., BUTTYÁN L. (2019): *Towards Detecting Trigger-Based Behavior in Binaries: Uncovering the Correct Environment*. Software Engineering and Formal Methods (SEFM 2019). Lecture Notes in Computer Science, vol 11724. Springer, Cham.

Okos mérők biztonsági tesztelése

Hallgató: Dömötör Szilárd

Mentor: Dr. Buttyán Levente

Az Internet of Things egyik célja, hogy a beágyazott eszközöket közvetlenül integrálja az Internetbe. Ennek különböző gazdasági előnyei, produktivitást és hatékonyságot növelő hatásai vannak, azonban az IoT eszközöket gyakran megfelelő biztonsági megfontolások nélkül fejlesztik, többek között az eszközök korlátozott hardveres erőforrásai miatt. A technológia elterjedésének egyik legnagyobb visszatartó ereje a biztonsági megoldások hiánya, ami gazdasági és fizikai veszteségeket okozhat.

Munkánk során először irodalomkutatást végeztünk beágyazott IoT eszközök biztonsági tesztelésének módszereiről, eszközeiről, majd ezt felhasználva sérülékenységeket kerestünk és használtunk ki vízfogyasztás mérésére használt okos mérő eszközökön. Az irodalomkutatás során általános firmware analízis technikákat, eszközöket, firmware letöltéséhez használható soros átviteli protokollokat, módszereket vizsgáltunk. A sérülékenység vizsgálathoz rendelkezésünkre állt két okos mérő, egy programozóegység, egy mérési adatok aggregálását végző eszköz és az ezekhez tartozó különböző kliensprogramok.

A vizsgálat során ellenőriztük a behatolás detektáló technológiák alkalmazását, mivel egy támadó közvetlen fizikai hozzáféréssel rendelkezhet az eszközökhöz. Az egyik mérőn nyitott JTAG debug portot találtunk, aminek használatával az eszköz firmware-ének letöltése lehetséges, ehhez kerestünk az eszköz mikrokontrollerével kompatibilis JTAG debuggert. Az eszköz teljes memóriatartalmát kinyertük, és elemeztük különböző bináris analízis eszközökkel. Kódvisszafejtést végeztünk a GHIDRA eszközzel, ehhez azonosítottuk és konfiguráltuk a memória régiókat a mikrokontroller datasheet alapján. Visszafejtettük az eszközök infravörös protokollon történő kommunikációját, és egy Man-in-the-Middle beékelődéses támadást alkalmaztunk ellene. Kriptográfiai megoldásokat elemeztük az eszköz által használt protokollokon, így azonosítottuk a protokollokon használt titkosítási kulcsokat. A mérő és programozóegység Bluetooth kommunikációját is visszafejtettük, majd implementáltunk egy Python scriptet a programozóegység instrumentálására, ami így a gyári szoftver nélkül használható. A protokollüzenetek manipulálásával és visszajátzásával az okosmérő konfigurációját, mérési értékeit megváltoztató támadást teszteltük. Megvizsgáltuk az aggregátor eszköz által gyűjtött adatok továbbítási módszerét és a kliens szoftverrel való hálózati kommunikációját is, ami titkosítatlanul viszt hitelesítési adatokat. Mindezzel párhuzamosan a kliens alkalmazások visszafejtését is elvégeztük, az ebből származó információk kiegészítették és megerősítették a firmware analízis eredményeinket.

Malware minták hasonlóságának vizsgálata

Hallgató: Tamás Csongor

Mentor: Dr. Buttyán Levente

Az internet használatának elterjedésével és a cégek digitalizációjával egyre gyakoribbá válnak a malware támadások. A készítőik a hatékonyság maximalizálása érdekében egy forráskódból sok végrehajtható mintát generálnak. Ezen folyamat során kisebb-nagyobb változtatásokat eszközölnek a fordításban, hogy azonos funkcionalitású, de mégis minél inkább eltérő variánsokat hozzanak létre.

Az antivírus szoftverek túlnyomó részt automatikusan generált szabályok használatával igyekeznek felismerni a kártékony alkalmazásokat. A szabályok eredményességét az elérhető mintahalmaz mérete és minősége határozza meg.

Munkánk során az volt a célunk, hogy egy malware mintákat tartalmazó adatbázisból olyan részhalmazokat tudjunk automatizáltan kiválogatni, mely részhalmazok egymáshoz hasonló mintákat tartalmaznak. Mivel a módszernek gyorsan kell működnie, dinamikus analízisre nincsen lehetőség, így a kutatás kizárólag automatikus módszerek vizsgálatát foglalta magában. A munka során számos elérhető statikus módszert megvizsgáltunk, ezek közül is főként a Locality Sensitive Hash algoritmusokat. Elemeztük az LZJD, SSDEEP, SDHASH és TLSH algoritmusok működését és összevetettük egyéb kutatásokban vizsgált módszerekkel. Megállapítottuk, hogy az ipari sztenderddé vált SSDEEP kevésbé alkalmazható malware klasszifikációs feladatok megoldására. Legjobban a TLSH nevű fuzzy hash algoritmus teljesített. A módszer alkalmazhatóságának bizonyítására implementálásra került egy keresési eljárás mely során egy több mint 300 000 000 mintát tartalmazó adatbázisból egy percen belül minden, egy keresési halmazhoz hasonló minta kiválogatható [1, 2]. A módszer számos új kutatásnak szolgálhat alapul, többek között lehetőséget ad más hasonlósági módszerek tesztelésére alkalmas malware adathalmaz létrehozására.

Referencia

- [1] CSONGOR Tamás: *Examples of Locality Sensitive Hashing and their Usage for Malware Classification*. Hacktivity 2018. (konferencia-előadás)
- [2] CSONGOR Tamás: *Method for similarity searching in large malware repository*. BSc Thesis, BME, December 2018.

Automatizált protokoll visszafejtés a hálózati forgalom statikus elemzésével

Hallgató: Ládi Gergő

Mentor: Izsó Tamás

Napjainkban számtalan olyan rendszert használunk, amelyek más rendszerekkel kommunikálnak. Ezen kommunikációk szabályait az úgynevezett protokollok írják le. A protokollok specifikációjában megtalálhatóak az elképzelhető (szabályos) üzenettípusok (azaz az egyes üzenetek formátuma és az üzeneteken belül az egyes bájtok értelmezése, jelentése), valamint az üzenetek lehetséges sorrendje (mely üzenet után mely más üzenetek következhetnek). Amennyiben birtokában vagyunk egy protokoll specifikációjának, úgy képesek vagyunk megérteni két (vagy több) fél kommunikációját. Ez segítségünkre lehet botnetek elemzésében, behatolásvédelmi és anomáliadetektációs rendszerek készítésében, valamint meglévő protokollimplementációk megfelelőségi és biztonsági tesztelésében. Sajnos sok számunkra érdekes protokoll zárt, azaz specifikációja nem ismert, vagy legalábbis nem elérhető a nyilvánosság számára.

A protokoll visszafejtés különböző módszerek együttese, melyek alkalmazásával (közelítőleges) specifikációkat tudunk készíteni zárt protokollokhoz. A kézzel történő visszafejtés fáradságos és folyamatos figyelmet igényel, így már számos kutatás foglalkozott a folyamat legalább részleges automatizálásán. A kutatások egyik csoportja egy konkrét implementáció gépi kódja alapján próbálja meg rekonstruálni a specifikációt, viszont az esetek egy részében nem férhető hozzá az implementáció sem, vagy esetleg a licencfeltételek nem engedik meg annak ilyen jellegű elemzését.

Jelen kutatás keretein belül kifejlesztettünk és publikáltunk egy kizárólagosan rögzített hálózati forgalom elemzésén alapuló protokoll visszafejtő algoritmust [1]. Az elmúlt évben ezen algoritmus továbbfejlesztésével foglalkoztunk. Az algoritmus új változata [2] képes több különböző (de ugyanazon protokollhoz tartozó) adatfolyamból készült modell összefésülésére, ezáltal pontosabb üzenetformátum-specifikáció visszafejtésére. Az algoritmus új variánsa így képes olyan összefüggések és mezőtípusok megtalálására is, amelyeket egyetlen adatfolyam vizsgálatával nem volna lehetséges felismerni. A következőkben a protokollok automatáinak (azaz az üzenetek lehetséges sorrendjeinek) visszafejtését tűztük ki célul.

Referencia

- [1] G. LÁDI, L. BUTTYÁN, T. HOLCZER: *Message Format and Field Semantics Inference for Binary Protocols Using Recorded Network Traffic*. IEEE Conference on Software, Telecommunications and Computer Networks (SoftCom), Croatia, September 2018.
- [2] G. LÁDI, L. BUTTYÁN, T. HOLCZER: *GrAMeFFSI: Graph Analysis Based Message Format and Field Semantics Inference*. Submitted for journal publication.

Konvergens IoT megoldások – Egységes szoftverfejlesztési módszertan

A kutatási program célja egységes szoftverfejlesztési módszertant biztosító elemek és folyamatok kutatása, azonosítása, kidolgozása és alkalmazása, melyek révén a konvergens szoftver megoldások irányába teszünk lépéseket.

Az IKT meghatározó, horizontális szereppel bír, az IKT felhordó ereje felelős a más szakterületek versenyképességének biztosításáért. A kutatási program eredményei olyan építőelemeket jelentenek, melyek hatékonyan támogatják a fenntartható tervezési, fejlesztési és karbantartási folyamatokat.

Mind a hardver infrastruktúra, mind pedig a szoftverrendszerek és alkalmazások képessége folyamatosan fejlődik, megfelelő alapot adnak a hatékony felhasználásra. A következő jelentős hozzáadott érték a kutatási és fejlesztési képességeinkben és kapacitásainkban jelenik meg. Ezért a következő lépés azon képesség formálása, megfelelő módszerek kidolgozásával, melynek révén a valódi megoldásokat fenntartható módon tudjuk kifejleszteni és a szakterületek számára szolgáltatni.

Convergent IoT Solutions – Unified Software Development Methodology

With the current program, our research team is formed to address a unified software development methodology with the goal to step forward to convergent software solutions.

ICT has a determining position and has a horizontal role, i.e., responsible for making other domains competitive. We are about to contribute to the building blocks that enable us to reach a methodology labeled with sustainable design, development, and maintenance. Therefore, we targeted to work on a Unified Software Development Methodology in order to realize convergent software solutions.

These points are examples to underpin, that both hardware infrastructure and software conditions are developing, they are appropriate to utilize them, and our further added value should be in our research and engineering capacities. Our next step is the capability that enables us to realize real solutions in a sustainable way and provide real value for the affected domains.



Dr. Lengyel László

Automatizálási és Alkalmazott Informatikai Tanszék
Department of Automation and Applied Informatics
lengyel@aut.bme.hu

Az Ipar 4.0 és Ipari IoT platformok megismerése

Hallgató: Batta Tamás

Mentor: Dr. Lengyel László

Manapság az IoT virágkorát éli és egyre nagyobb igény keletkezik olyan szolgáltatásokra, amelyek nagy hangsúlyt fektetnek az imént említett technológiára. Nem véletlen, hiszen az IoT ígérete, többek között, az egyes környezetek, tárgyak megfigyelése, azok adatainak kiolvasása és mentése, azért, hogy azokból tette fogható információt legyünk képesek előállítani, azaz, olyan információra tegyünk szert, amellyel valamilyen formában értéket teremthetünk. Az technológia alkalmazásának nincs iparági korlátja, így szinte minden területen alkalmazható, így például az iparban is. Az IoT nagy hangsúlyt kap az Ipar 4.0-ban, avagy a negyedik ipari forradalomban, ahol a rengeteg megoldandó feladatot, mint például „predictive maintenance”-t, automatizálást stb., elképzelhetetlen megoldani a technológia felhasználása nélkül.

Egy IoT rendszer kiépítése meglehetősen komplex feladat, olyannyira, hogy a dokumentum írásának idején már nem szokás kialakítani megfelelő IoT platform kiválasztása nélkül, amely megoldja többek között a szenzorokból jövő adatok fogadását és azok (ideiglenes) tárolását, sőt mára már olyan funkciókkal is kibővültek, mint pl.: saját felület kialakítása, adatok, avagy információ megjelenítésére.

Az ösztöndíj ideje alatt, megismerkedtem az Ipar 4.0 világával, az ottani kihívásokkal, követelményekkel, és alapvetően két, az előbbieken ismertetett, kifejezetten ipari felhasználásra készült IoT platformmal is, a PTC – Thingworx, illetve a Siemens – Mindsphere-el. Ez idő alatt megismertem a két rendszer alapvető működését, hogy milyen modell alapján közelítik meg az Ipari IoT világában lévő problémákat. Megtanultam, hogyan telepíthetek ki eszközöket különböző környezetekbe, hogyan kell tervezni egy alap IoT architektúrát, hogyan menthetek le adatokat az egyes platformon, illetve azokat hogyan jeleníthetem meg üzleti felhasználásra. A platformokon való munka mellett továbbá megtanultam, hogy hogyan és miként érdemes egy eszközt elkészíteni, amely majd adatokat lesz képes szolgáltatni az egyes platformok felé. Végezetül, hogy a megszerzett tudást gyakorlatba is áthelyezzem, több proof-of-concept-et is elkészítettem, hogy lássam a rendszer éles működését.

IoT a gyógyszergyártásban

Hallgató: Köte Ákos

Mentorok: Dr. Csorba Kristóf, Dr. Lengyel László

A PharmaTech laboratóriumban egyik célunk egy olyan dinamikus képelemzésen alapuló rendszer megalkotása, aminek segítségével nagy pontossággal lesz mérhető a gyógyszerhatóanyagok tömegárama alacsony adagolási sebességek mellett is. A képekből kinyerhető az azokon látható szemcsék fényessége, körülhatárolása, majd tömegárama. Implementáltuk a mérlegről történő adatbeolvasást, az adagoló szabályozott irányítását, valamint a kamera vezérelt elsütését és beolvasását. Az adagoló számítógépről történő vezérléséhez szükséges egy arra alkalmas vezérlőegység fejlesztése, illetve az azzal történő kommunikáció kialakítása. Másik célunk még a rendszer modellezése, valamint annak kamera alapú visszacsatolós szabályozása, aminek révén a hatóanyag tömegáram 5%-os hibahatáron belül legyen. További célunk a gyógyszer kristályosítási folyamatok képszegmentációval történő in-line ellenőrzése, a kristályok morfológiájából következtethetünk a folyamat eredményességére.

A szoftver futási idejének biztosítása kritikus fontosságú, ugyanis egységnyi tömegű szemcse kétszer gyorsabb képsűrűséggel kétszer több összfényességet eredményez. Ezért szükséges volt az eleinte a felhasználói felülettel együtt C#-ban fejlesztett képelemző modul C++-ba történő átírása, amivel stabilabb futási idők érhetők el. Ilyen módon a megjelenítő felület elválasztásra került a feldolgozóegységtől.

Az általunk használt kamerák nagy sebességű (kb. 200 FPS) felvételeket készítenek, ezért a képkockák tárolásához egy gyors, jó tömörítésű és minőségű módszerre van szükségünk. Több különböző kódoló/dekódoló kipróbálása után a H.264 nevű tömörítő eljárás tűnt a legideálisabbnak. A képkockákból az FFMPEG szoftver segítségével készítjük a videókat, amiket utólagos elemzéseinkhez használjuk.

A visszacsatolós szabályozás tervezése során az adagoló sebessége és a por tömegárama közti összefüggést vizsgáltuk. A rendszer (ARX) modellezéséhez az úgrázavarokra adott válaszokat vizsgáltuk, és elkészítettük a tömegáramhoz tartozó PID szabályozót.

Hálózati forgalom vizsgálatára épülő adatvédelem az okosotthonban

Hallgatók: Dominguez Zoltán, Kováts Levente

Mentor: Dr. Gulyás Gábor György

Az IoT térhódításával párhuzamosan az okos eszközök megjelentek a háztartásokban is. Ezek elsősorban szenzorok és internetkapcsolat segítségével igyekeznek többet nyújtani, mint „hagyományos” társaik. Ennek azonban az ára, hogy az otthoni érzékeny adatok kiszivároghatnak. Például az okostévéken már megjelentek a fogyasztási szokások gyűjtésre szakosodott trackerek (amelyek működésüket tekintve hasonlóak webes társaikhoz), illetve szintén prominens példa a kompromittálódott eszközök (pl. kamerák, okostelefonok) által okozott szívargási problémák és visszaélések.

Ezen eszközök ökoszisztémája már most túl komplex ahhoz, hogy felhasználók ne legyenek képesek átlátni a potenciális adatvédelmi kockázatokat, sem azokat befolyásolni. Erre megoldást kínálhat egy olyan hálózati eszköz, amely a hálózati forgalom alapján képes releváns kimutatásokat készíteni, és lehetőséget ad a nem kívánt kommunikációk blokkolására. Munkánkban egy ilyen hálózati szűrő alkalmazáson dolgoztunk, Python nyelven Raspberry Pi platformra. Mivel a teljes hálózati forgalom mély elemzése (deep packet inspection) rendkívül erőforrás igényes, olyan megoldás mellett döntöttünk, amely elsősorban a DNS lekérdezéseken keresztül végzi tevékenységét.

A projektben elkészült megoldás két fő komponensből állt, az egyik a DNS lekérdezéseket kiszolgáló és kezelő mag, a másik az ehhez kapcsolódó webes felületet és API-kat szolgáltató rész. A mag a webes felület nélkül is működőképes. A mag rész nagy sebességgel kiszolgálja a kéréseket, azokat naplózza is, és lehetővé teszi modulok közbeiktatását a kiszolgálási folyamatba, például statisztikai feldolgozást és a szokásostól eltérő viselkedési minták kiszűrését. A jelenlegi modulok szabályalapúak, de az elkészült struktúra lehetővé teszi gépi tanulás alapú modul hatékony használatát is.

Felmértük a reklám, tracker blokkolási listák helyzetét. Azt találtuk, hogy ezek a legtöbbször redundánsak, készítőik sokszor mélyebb megfontolás nélkül másolnak egymásról. Azonban a legeredetibb listákra leszűkítve a mezőnyt mértük azok hatékonyságát is, amely előre jelzi egy hálózati szűrő eszköz hatékonyságát is. Ehhez a Trunco weboldal látogatottsági listát és az OpenWPM eszközt használtuk fel. Eredményeink alapján egy egyszerű listás eszköz már hatékony alapvédelmet nyújthat.

Referencia

- [1] Kováts LEVENTE: *Jelentés a kutatási/fejlesztési tevékenységekről 2019. első negyedéről*. Kutatási jelentés. (2019. május 24.)

Anonimizálás folyamatosan érkező adatokban

A 2018 május 25-étől hatályos GDPR (General Data Protection Regulation) előírással új korszak kezdődött az adatot kezelő vállalatoknál. Egy vállalat működése során természetesen adatot rögzít. Ennek egy része az üzlet mindennapi működéséhez szükséges; más része pedig az üzlet fejlesztése céljából kerül rögzítésre. Egy vállalat sok adminisztrációs és jogi teherrel tud mentesülni, ha az adatok egy részét, legalább azt a részét, ahol ez nem feltétlenül szükséges, személyes adatok nélkül tárolja.

Kutatásunkban olyan GDPR-kompatibilis anonimizáló megoldásokat fejlesztünk ki, és fejlesztünk tovább, amelyek az adatkezelőt minél jobban mentesítik a személyes adatok kezelésének problémája alól. Célunk olyan módszerek kidolgozása, amelyek nagy adathalmazokon is oly módon tudnak működni, hogy azt automatikusan, az adatrögzítés helyén anonimizálják, illetve amennyiben ez nem lehetséges, csak teljesen automatizáltan feldolgozva a beérkezés pillanatában törölnek belőle minden nem lényeges adatot.

Anonymization of stream data

The General Data Protection Regulation (GDPR), which came into force on May 25, 2018, invoked a new era for companies managing data. A lot of data is recorded during the operation of any company. Part of this is necessary for the day-to-day operation of the business; the other part is recorded for the purpose of business development. A company can be relieved of administrative and legal burdens if at least part of the data is stored without personal data.

In our research, we are developing GDPR-compliant data anonymization methods that lift the burden of the need to handle privacy-sensitive data. Our goal is to offer techniques that work on arbitrarily large data sets and automatically anonymize the data at the place of its generation before the data is sent to central storage. In case this is not feasible, we are looking at fully automated workflows that strip the incoming data of any unwanted and privacy-sensitive details.



Dr. Dudás Ákos

Automatizálási és Alkalmazott Informatikai Tanszék
Department of Automation and Applied Informatics
Dudas.Akos@aut.bme.hu

Anonimizálás a Differential Privacy modell és a k-anonimitás kombinálásával a GDPR-nak való megfelelés érdekében

Hallgatók: Majkut Kristóf, Soós Sarolta

Mentor: Dr. Dudás Ákos

A 2018-as év a vállalatok számára a GDPR bevezetésére való felkészülést jelentette. Bár az adatvédelmi rendeletet 2016-ban fogadták el, csak tavaly lépett hatályba. A GDPR értelmezése és megfelelő alkalmazása még napjainkban is kihívás elé állítja a cégeket, illetve nagy terhet ró mindenkire, akik valamilyen módon adatokkal dolgoznak. A gyakorlati életben a probléma abban rejlik, hogy a rendelet ugyan szigorúan szabályozza a különböző adatok kezelését, nem ad azonban pontos iránymutatást arra vonatkozóan, hogy miként kell az említett szabályokat betartani. A GDPR tehát jogi előírás, és nem technológiai, így a vállalatok arra kényszerülnek, hogy saját maguk alakítsák ki a megfeleléshez szükséges folyamataikat.

Kutatásunk során egy olyan anonimizáló rendszerrel ismerkedtünk meg, mely képes akár folyamatosan érkező adatok befogadására és anonimizálására [1]. Ezen architektúrára alapozva célunk egy olyan GDPR-kompatibilis megoldás bemutatása, melynek felhasználásával az adatkezelők automatikusan teljesítik a rendelet követelményeit. Az általunk javasolt módszer kiindulópontja az irodalomból jól ismert k-anonimitás. Mivel az anonimizált adatok nem esnek a GDPR hatálya alá, ha mindent az adatrögzítés helyén anonimizálunk [2], s később is csak anonim formában tároljuk az információkat, mentesülhetünk a GDPR-előírások alól.

Ugyan a GDPR körüli nehézségeket kiküszöböli, az anonimizálás önmagában mégsem nyújt 100%-os adatbiztonságot. A sérülékenységeket kihasználva egy k-anonim adathalmazban szereplő személyek is beazonosíthatók. Ennek okán kutatásunk másik célja, hogy a javasolt algoritmust minimális változtatásokkal kiterjesszük úgy, hogy a Differential Privacy modell által támasztott követelményeknek is megfeleljen.

Referenciák

- [1] MARTINEK Vilmos: *Anonimizálás big data környezetben*. Diplomaterv, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2018.
- [2] MAJKUT, Kristóf, DUDÁS, Ákos: *K-anonymity-based client-side anonymization*. István, Vajk; Dmitriy, Dunaev (szerk.) *Proceedings of the Automation and Applied Computer Science Workshop 2019* (2019) pp. 109-118.

Big Data SQL végrehajtó automatikus választása machine learning alapokon

Hallgató: Maidics Barnabás

Mentor: Dr. Dudás Ákos

A nagyméretű adathalmazok feldolgozása mindig is fontos szerepköre volt az informatikai rendszereknek. Jelenleg amennyiben a felhasználó egy lekérdezést szeretne futtatni egy ilyen ún. big data adathalmazon, akkor több eszköz közül is választhat. Ezek közül kettő legelterjedtebb, nyílt forráskódú megoldás az Apache Hive [1] és Apache Impala [2]. A két megoldást nagyon gyakran párhuzamosan használják egy valós környezetben. Bizonyos esetekben a Hive, másokban az Impala tud gyorsabb végrehajtási időt elérni.

Jelen pillanatban a felhasználó feladata eldönteni, hogy melyiken fogja futtatni az adott lekérdezését. Ez a feladat viszont meglehetősen nehéz, összetett. Rengeteg paramétertől függ a végrehajtási idő: lekérdezésben szereplő táblák mérete, azok particionálása, az SQL lekérdezésben lévő illesztések, csoportosítások száma, az egy-másba ágyazott allekérdezések száma. Ezekon kívül feltételezhető, hogy sok más, ismeretlen, manuálisan nehezen felderíthető tényezőtől. Projektem, kutatásom célja, hogy megkönnyítsük a felhasználók feladatát azzal, hogy automatikusan kiválasztjuk az optimális végrehajtó egységet az adott lekérdezéshez. Ezzel összességében a felhasználói élményt és a rendszerünk áteresztőképességét javíthatjuk.

A projektem neve QuDi (Query Distributor), amely egy proxy réteggént működik a felhasználói réteg és az Apache Hive, Impala között. A bejövő lekérdezést elemzi, azokból különböző változókat gyűjt (pl.: particiók, illesztések, csoportosítások száma, a lekérdezésben szereplő táblák mérete stb.), majd ezek alapján egy machine learning modellt felhasználva eldönti, hogy melyik végrehajtó egységen érdemes az adott lekérdezést futtatni a lehető leggyorsabb válaszidő érdekében.

Referenciák

- [1] Ashish THUSOO, Joydeep SEN SARMA, Namit JAIN, Zheng SHAO, Prasad CHAKKA, Ning ZHANG, Suresh ANTHONY, Hao LIU and Raghotham MURTHY: *Hive – a petabyte scaledata warehouse using Hadoop*. 2010.
- [2] Cloudera Inc. Impala: https://docs.cloudera.com/documentation/enterprise/5-8-x/topics/impala_concepts.html.

Mesterséges intelligencia alkalmazása GDPR-nak megfelelő adatfeldolgozás során

Hallgató: Szondy Dániel

Mentor: Dr. Dudás Ákos

A jelenleg hatályban levő GDPR szabályozás szigorúan meghatározza adatok kezelésének módját. Munkámban *célhoz kötöttség* elvének betartása és adatfeldolgozók szerepének minimalizálása érdekében keresek mesterséges intelligenciával támogatott megoldásokat szöveges adatok (például levelek, embertől érkező, vagy generált üzenetek) esetén a tartalom feldolgozására. Az alapvető nehézség, hogy a szöveg tartalmáról nem tudunk egyértelműen nyilatkozni, így nem zárható ki, hogy érzékeny adatokat is tartalmaz. Az adatkezelési incidensek kockázatát szeretnénk minimalizálni, így szükséges a lehető legnagyobb mértékben gépi feldolgozással végezni a szöveg értelmezését.

Legyen a konkrét példánk egy e-mailben érkezett megrendelés és ennek feldolgozása. A levél tartalmazhatja a megrendelés részleteit, melyre minden esetben szükség van, a szövegből mindenképp ki kell nyerni. Másik kategóriába kerülhetnek az ennél érzékenyebb, de szükséges adatok, mint számlázási adatok, lakcím és más hasonló tartalom. Ezen adatok szintén szükségesek, ugyanakkor sok esetben nem kívánatos, hogy ezen adatokhoz „nem megbízható” módon lehessen hozzáférni.

Egy utolsó kategóriába tarthatnának olyan adatok melyeket, habár a levél küldő fél megadott, de a feladat szempontjából nincs relevanciája. Adatkezelés szempontjából ezek érdekesek, mivel lehetőleg el kell kerülni feldolgozásukat, rögzítésüket. Ezek az adatok olyan helyzetben kerülhetnek elő, amikor a levél nem célzottan a kezelőnek szól, csak részlegesen vesz részt a folyamatban, mint például foglalásról készít naptári értesítőt, amilyen esetben a tartalmazott adatok nagy része nem releváns.

A probléma nagyban hasonlít a „Task-Oriented Dialogue Modelling” (vagy „Goal-Oriented”) feladatokra hasonlít [1], ugyanakkor sok esetben hosszabb szövegeket feltételez, mint az ilyen esetekben jellemző néhány mondatból álló szövegrészletek. Munkánkban ezen megoldások alkalmazhatóságát is vizsgáljuk.

Referencia

- [1] Paweł BUDZIANOWSKI, Tsung-Hsien WEN, Bo-Hsiang TSENG, Iñigo CASANUEVA, Stefan ULTES, Osman RAMADAN and Milica GASIĆ: *MultiWOZ – A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling*. 2018.

Szolgáltatásbiztonság garantálása kritikus CPS rendszerekben

A kiberfizikai (CPS) rendszerek a (termelési, az emberi okos*, stb.) környezet folyamatait forradalmian magasabb minőségben szabályozzák a lokális beágyazott rendszereket az Interneten elérhető felhő számítási kapacitását és algoritmikus intelligenciáját kombinálva. A kritikus CPS-ek informatikai hibáit a fizikai környezet akár katasztrófálissá is felerősítheti.

A kutatás célja a szabványosodó, Edge alapú CPS architektúrák feletti hibatűrő minták modell alapú kidolgozása, és benchmark mérésekből vagy működési naplóból teljesítőképességi és szolgáltatásbiztonsági modellek és analízis megoldások kidolgozása. Ehhez a gépi tanulás eszközeit is integráljuk.

A demonstrátorok a hibatűrő megoldást az OMG Distributed Data Service szabványos implementációja felett mutatják be. Új elemeként a hibák elhárítására illetve változó igények kiszolgálására dinamikusan adaptív architektúrákat használunk. A kritikusság helyes kezelését a teljesítőképesség formális elemzésével igazolják.

Dependability assurance in critical CPS systems

Cyber-physical systems (CPS) provide a revolutionary new quality in controlling the processes in the (human smart-*, production, etc.) environment. They combine the local embedded systems with cloud resources and algorithmic intelligence via the internet. However, the physical environment may amplify computing faults to a catastrophic level.

Our research focuses on fault-tolerant design patterns over edge-computing based CPS architectures. Performability and dependability model identification and validation, starting from benchmark experiments or operational logs supported by machine-learning, is the primary paradigm used.

The demonstrators present the novel concepts for fault-tolerance and resiliency over the OMG industrial standard Distributed Data Service. In particular, a novel mode of use of dynamic adaptive architectures serves for fault mitigation and adaptation to changes in the requirements. Prototypes for formal analysis of performability characteristics are in their pilot phases.



Dr. Pataricza András

Méréstechnika és Információs Rendszerek Tanszék
Department of Measurement and Information Systems
pataric@mit.bme.hu

Modellalkotás mérési adatokból

Hallgató: Földvári András, Hodvagner Szilvia, Burján Dezső

Mentor: Dr. Pataricza András, dr. Gönczy László

A CPS rendszerek célja a fizikai világ intelligens megfigyelése, illetve irányítása. A CPS alkalmazásoknál kritikus az extrafunkcionális kritériumok kielégítése (időbeliség, át-bocsátóképesség, megbízhatóság, rendelkezésre állás stb.).

Tervezéskor az infrastruktúra méretezése és a működés alatt a futásidejű dinamikus erőforrás allokáció egyaránt igényli a követelményeket, a rendszer architektúrát és a mért metrikákat egységesítő modell megalkotását a benchmark mérések vagy operációs naplók feldolgozásával végzett rendszeridentifikációval.

A CPS rendszerek üzemmodjai és azon belüli karakterisztikái az erősen nemlineáris és sok-paraméteres viselkedésük miatt nem írhatóak le analitikus módszerekkel kellő valósághűséggel. Egyre inkább követelmény a modell skálázhatósága is, azaz egyazon algoritmus különböző méretű feladatokra történő alkalmazásánál a modell adaptálása a feladat aktuális méretéhez. Ennek legcélszerűbb módja a hibrid modellezés, mely a méretfüggetlen viselkedési tartományokat diszkrét értékekkel jellemzi és a skálázás az ilyen kvalitatív tartományok határainak megválasztásával történik.

A rendszeridentifikáció egyik leghatékonyabb módszere a vizuális interaktív fel-táró adatanalízis (EDA), amelynél különböző grafikonokon térben és időben megjelenített mérési adatok közötti összefüggést célterületi szakértő vizsgálja a rendszerről, illetve a követelményekről való háttérismerete felhasználásával.

Meglevő matematikai vizualizációs csomagok fölé olyan keretrendszert alkotunk, amely támogatja a kvalitatív tartományok meghatározását (pl. válaszdő alacsony-közepes-magas) és döntéstámogatói módszereket (pl. véletlen erdő, XGBoost) használva a célváltozó értékeinek meghatározását.

Az EDA folyamatát foglalja szervezett keretbe az a újszerű, a technikai hibaanalízis egyik bevált módszerét, az integrált diagnosztikát a modelépítési folyamatra adaptáló megközelítésünk, mely garantálja a részletekből épített modell konzisztenciáját.

Komplex rendszerekre a heurisztikus megközelítés nehézkes. A mesterséges intelligencia egyik módszerét, a kvalitatív tartomány feletti gépi tanulást (induktív következtetés) alkalmaztuk az előzetes ismereteket tudásgráfban reprezentálva az identifikációs modell automatikus származtatására a mérési adatokból.

Dinamikus rekonfiguráció edge alapú kritikus CPS rendszerekben

Hallgatók: Szalontai Jenő, Pethő Dániel, Szabó Richárd

Mentor: Dr. Kocsis Imre

Az edge és „fog” számítástechnika a felhő számítástechnikát kiterjeszti a „terep” irányába; míg az edge computing a „felhő és az internet határán” (pl. közvetlenül 5G bázisállomások mellett) hoz létre felhő-szerűen igénybe vehető szolgáltatásokat, addig a „fog” számítástechnika a felhő szolgáltatási modell alkalmazását akár a terepi, illetve végfelhasználói eszközökön is lehetővé teszi.

Ezen rendszerek a „terep felé haladva” növelik az alkalmazások lehetséges lokalitását (pl. szenzoroknál és beavatkozóknál) és csökkentik a késleltetést (kemény valós idejű alkalmazások). Az alkalmazások funkcionális architektúrája terepi, edge és távoli felhő szolgáltatásokra is allokalhatóvá válva az allokáció dinamikusan módosítható. Ez főként alkalmazások ellenállóképességének (resilience) biztosítása mellett szempontjából fontos.

Az újrakonfiguráció futásidejű tervezésének fő kihívása ma már a tudásreprezentáció. Egy adott alkalmazás funkcionális elemei jellemzően többféleképpen is megvalósíthatóak, de a funkcionálisan ekvivalens megoldások felismerése igényli a szenzor- és beavatkozó-képességek szemantikus reprezentációját.

Kidolgoztuk az általános edge CPS újrakonfigurációs probléma egy reprezentatív esetét támogató megközelítést. A szenzor-adatfolyamok feldolgozásának funkcionális specifikációját a W3C Semantic Sensor Network ontológiája alapján végezzük, a részfunkciók megvalósítását Kubernetes feletti mikroszolgáltatásként telepítve. Az alkalmazások immutábilis „vázát” a szabványos OMG DDS platform adja, mely a publish-subscribe elvű rendszerintegráció extrafunkcionális követelményeinek teljesülését finom felbontással biztosítja.

Az újrakonfiguráció tervezéséhez szemantikus következtetést támogató (triple store) adatbázis követi folyamatosan nyomon a rendszerelemek konfigurációját és állapotát. A DDS csatornák „között” elhelyezkedő funkciók szemantikusan ekvivalens megvalósítási lehetőségeinek feltérképezésére a formális fogalomelemzést (Formal Concept Analysis, FCA) alkalmazzuk.

Kritikus CPS rendszerek formális módszerekkel támogatott analízise

Hallgatók: Bajczi Levente, Dobos-Kovács Mihály

Mentor: Dr. Vörös András

A kritikus CPS rendszerekben a fizikai rendszert valós időben vezérlő beágyazott eszközök egyre bonyolultabbak. Szolgáltatásbiztonságuk garantálásában fontos a hibákat a fejlesztés korai fázisaiban felfedő formális verifikáció.

A többmagos processzorok terjedésével egyre nagyobb az igény a számítási kapacitásuk kihasználására. A párhuzamos programok azonban időben nem-determinisztikusak, így fejlesztésük még az egyszálas programozásnál is több buktatót rejt. A konkurens lefutások közötti versenyhelyzetek miatt a lehetséges lefutások száma egyszerű esetben is hatalmas.

A többmagos processzorok legfontosabb tulajdonsága a magok közötti szinkronizációt befolyásoló memória modell, ugyanis ennek hibáit a szoftverben kell kivédeni. A kutatás során a többmagos processzorok memória modelljeit vizsgáló hardver verifikációs eszközöket is felhasználjuk a szoftver verifikáció során együttműködésük valósághű modellezésére.

Kutatásunk másik iránya az autóiipari beágyazott CPS rendszereket célozza meg a formális módszerek recizitását ötvözve a tesztelés hatékonyságával. Ezen algoritmusok az ellenpélda alapú absztrakció finomítás módszerét kombinálják a SAT és SMT megoldók hatékony interpoláns generálásával annak érdekében, hogy egy jó absztrakciót találjanak, amelyen sikeresen bizonyíthatóak a tulajdonságok. Mi ezt a reprezentációt használjuk fel a továbbiakban tesztgenerálási célra. A megközelítést integráljuk az ipari szabvány AUTOSAR környezetbe.

Demonstrátorok szolgáltatásbiztos CPS rendszerek bemutatására

Hallgatók: Szalontai Jenő, Szántó Benedek, Szabó Richárd, Kilián Csenge

Mentorok: Dr. Vörös András, Dr. Pataricza András, Dr. Kocsis Imre

A kutatások során fontosak az egyes új módszerek empirikus validációjára és bemutatására alkalmas demonstrátor rendszerek. Ennek érdekében több demonstrátor rendszert is megalkottunk, melyek közös jellemzője, hogy ipari szabvány, magas rendelkezésre állást és valós idejű beavatkozást lehetővé tevő DDS rendszerintegrációs köztes rétegre épülnek. A szabványos integrációs technika használata lehetővé teszi újabb eredmények (pl. intelligens adatfeldolgozás, bővebb információ fúzió stb.) szintre ráfordítás mentes bevonását is.

Az összes demonstrátor SCADA-szerű, de egyszerűen implementálható open source alapú „műszerezettség”, vezérlési és vizualizációs technika integrálásával teszi a szakterületi szakértők számára érthetővé az egyes megoldások nyújtotta szolgáltatás szintű előnyöket.

Munkánk során egyrészt befejeztünk egy üvegház demonstrátort, amely szemlélteti, hogy miként lehet egy meglévő fizikai infrastruktúrát annak alapvető módosítása nélkül pótlólagosan felokosítani a kézi vezérlést kiváltó korszerű, hatékony CPS-sé kiegészítő intelligens beavatkozással és külső adatforrásokból származó információ fúziójával.

Elkészült egy smart-grid rendszer váza is, amely az intelligens prediktív vezérlés mellett a valós idejű beavatkozást és hibatűré tervezési minták alkalmazását is szemlélteti. Ezen demonstrátor fejlesztése során újra felhasználtuk az előző, adaptív rendszerek tervezésével foglalkozó kutatásunk eredményeit is, hiszen a hibatűrés és átkonfigurálás fontos részét képezi a demonstrátornak.

A demonstrátorok illusztrálják a korszerű, olcsó, de teljesítőképes mini-számítástechnikán alapuló forradalmi változást is. A hatékony rendszerintegrációs platformok ugyanis képesek ilyen eszközökből is elosztott rendszerekben is magas fokon garantált szolgáltatás biztonságot garantálva extrém jó költséghatékonyságbeli generációs ugrást létrehozni a teljes rendszer intelligenciáját és szolgáltatásait tekintve.

Infokommunikációs hálózatok és kiberfizikai rendszerek (B pillér)

Infokommunikációs hálózatok és kiberfizikai rendszerek („B” pillér)

A pillér célja a kommunikációs rendszerek infrastrukturális feladataival és azok alkalmazásával kapcsolatos kutatási területek összefogása. A pillér a BME kutatási tevékenységéhez illeszkedve, hat tématerületet ölel fel, melyek változatosak, de nem függetlenek; kiegészítik és támogatják egymást. A kiadvány húsz hallgató munkáját mutatja be, akik a BME VIK öt tanszékén dolgozó kutatók irányításával dolgoztak. A közös munka a célcsoport és a mentorok számára is hasznos tudományos eredményeket hozott. A jövő kutatói generációjának támogatása mellett együttműködési lehetőséget teremtett a BME tanszékei és természetesen a BME, ELTE és PPKE munkatársai között.

A valós idejű CPS rendszerek téma a kiberfizikai rendszerek valós idejű futásához szükséges ütemezés és az óraszinkron kérdéskörrel foglalkozik. Alacsony árú fejlesztőkártyákon és open hardware alapon felépített mesteróra kifejlesztését végzik a támogatott hallgatók, ezek kívül foglalkoznak a kommunikációs hibák kezelésével és alkalmazási példaként oktatási intézményekben a levegő minőségének monitorozásával.

A felhő alapú hálózati megoldások tématerület a hálózat virtualizációjára helyezi a hangsúlyt. A szerver nélküli számítások során, virtuális infrastruktúra kiépítése és menedzselése nélkül vehetők igénybe szolgáltatások. A komplex hálózati szolgáltatások megvalósítása, a megoldások teljesítményének, optimalizálási lehetőségeinek vizsgálata fontos kutatási terület.

A számítási felhő és minősége tématerület a felhő infrastruktúra megvalósításával foglalkozik. A felhőmenedzsment rendszerhez egy új ütemezési algoritmus került kifejlesztésre, ami figyelembe veszi a korlátozott mintavételi gyakoriságból adódó pontatlanságokat.

A flexibilitásra és átkonfigurálhatóságra helyezi a hangsúlyt a **Szoftverrádió alapú IoT fejlesztőrendszer** téma, mely a vezeték nélküli infrastruktúra képességeit vizsgálja. Az alkalmazott platformmal a kifejlesztett módszerek teljesítőképességét a gyakorlatban mérésekkel lehet tesztelni. A kutatás kiemelt feladata a masszív több-antennás (massive MIMO) technológiák vizsgálata.

A időkritikus 5G hálózati infrastruktúrák a hozzáférési hálózat új irányait vizsgálja. Az 5G hálózatokban az optikai szegmens kiterjesztésére, a szabadtéri szakaszon való optikai megközelítéssel és a biztonságos kommunikációhoz szükséges kvantumos elvű véletlenszám generátor elméleti és gyakorlati kutatásával foglalkozik.

A digitális infrastruktúra szolgáltatási képességeinek fejlesztése és automatikus üzemeltetésének megalapozása téma a hálózati megoldások skálázhatóságára, flexibilitására, átprogramozhatóságára koncentrál. A rendkívüli nagyszámú és nagy sűrűségben jelenlévő, részben mozgó végpont hálózati kiszolgálása miatt az 5G rendszerek szolgáltatási képességeinek fejlesztése a célkitűzés.

Info-communication Systems and Cyber-Physical Systems („B” pillar)

The pillar deals with research areas related to the infrastructure of communication systems and their applications. In line with BME's research activities, it includes six subject areas that are diverse but not independent; they complement and support each other. Twenty students have worked under the guidance of researchers from the five departments of BME VIK. The project supports the future generation of researchers, and it creates an opportunity for collaboration between BME departments and, of course, BME, ELTE, and PPKE staff.

Real-time CPS systems topic deals with scheduling the real-time running of cyber-physical systems and the clock synchronization issue. Supported students develop low-cost, open hardware-based master classes, besides addressing communication errors and monitoring air quality in educational institutions as an example.

Cloud Networking Technologies topic focuses on network virtualization. Serverless computing offers services without the need to build and manage a virtual infrastructure. The implementation of complex network services, the investigation of the performance and optimization of the solutions are essential research areas.

Cloud Computing and its Quality theme addresses the implementation of cloud infrastructure. A new scheduling algorithm has been developed for the cloud management system, which takes into account inaccuracies due to the limited sampling frequency.

Flexibility and reconfigurability are emphasized in the **SDR-based IoT Prototyping Platform** theme, which explores the capabilities of wireless infrastructure. With the applied platform, the performance of the developed methods can be tested. The main task of the research is the investigation of massive MIMO technologies.

5G time critical architectures are exploring new directions in the access network. In 5G networks, it deals with the theoretical and practical research of optical.-wireless approach, and the quantum random number generators required for secure communication.

Methods to Enhance the Service Capability of Digital Infrastructure topic establishes the new challenges and issues of 5G networks, like scalability, flexibility, reliability, and latency.



Gerhátné Dr. Udvary Eszter

Szélessávú Hírközlés és Villamosságtan Tanszék

Department of Broadband Infocommunication

and Electromagnetic Theory

udvary@hvt.bme.hu

Valós idejű CPS rendszerek

A Kiber-fizikai rendszerek (Cyber-Physical Systems, CPS) alkalmazások jelentős részében az aktivitásokat valós időben (RT-CPS) kell elvégezni. Ezért az IEEE 1588 óraszinkronizációs protokoll vagy annak variánsainak megvalósítása elkerülhetetlen ezekben a rendszerekben, ugyanakkor a piacon kapható ilyen eszközök magas árak és zárt megvalósításuk miatt korlátozottan alkalmazhatók kutatási és oktatási környezetben. A probléma megoldására egy alacsony árú, elérhető fejlesztőkártyákon és open hardware alapon felépített mesteróra kifejlesztését tűztük ki egyik feladatként. Ezen kívül kutatásaink kiterjednek a RT-CPS környezetben az elkerülhetetlen kommunikációs hibák kezelésére is, amelyek a CPS szinten adatvesztés formájában jelentkeznek a valós-idejű rendszerekben, mivel újraküldésre ezekben nincs idő a valós-idejű előírásokból következően. Harmadik kutatási területünk a levegő minőség monitorozása oktatási intézményekben, ahol a célunk egy olyan alacsony költségű rendszer kifejlesztése, amely lehetővé teszi a tanulást nehezítő, egészségtelen levegő jelzését előadótermekben és laboratóriumokban.

Real-Time CPS Systems

Most of the Cyber-Physical System (CPS) applications require that activities must be executed in real-time (RT-CPS). Therefore, the implementation of IEEE 1588 or its derivatives is unavoidable in such systems. However, the price and the closed hardware and software implementation of commercial IEEE 1588 compatible equipment make it very hard to use them in research and education. As a solution to the problem, we initiated the development of a low cost, publicly available development board, and open hardware and software paradigm based master clock. CPS systems must be ready to handle data loss, as it is impossible to resend information in real-time systems in case of communication problems due to the deadlines defining the systems, so our research is extended toward this direction also. Our third research area is air quality monitoring in educational facilities, for which we develop a low-cost monitoring system making possible to detect bad quality, unhealthy air in lecture halls and laboratories.



Dr. Kovácsházy Tamás

Méréstechnika és Információs Rendszerek Tanszék
Department of Measurement and Information Systems
khazy@mit.bme.hu

IEEE 1588 mesteróra fejlesztése kísérleti CPS rendszerekhez

Hallgató: Hollós Ádám Erik

Mentor: Dr. Kovácsházy Tamás

Kutatásom célja egy IEEE 1588 mesteróra létrehozása, mely GNSS szolgáltatások igénybevételel (pl. GPS, Galileo) képes magát a globális (UTC) időhöz szinkronizálni, majd ezután képes a helyi hálózat (LAN) többi óráját PTP (Precision Time Protocol, IEEE 1588) segítségével magához szinkronizálni. Fontos, hogy az elkészült eszköznek olcsónak kell lennie és az Open Hardware paradigmát kell követnie. Ugyanis, a cél az, hogy a PTP és a TSN (Time Sensitive Networking) kutatását és tanítását könnyítse meg az egyetemek számára is megengedhető árával.

A kutatásom során nagy hangsúly helyeződött az alkalmazott GNSS vevő kiválasztására, mivel a vevő jelenti a rendszer lelkét, hiszen ő mondja meg / számítja ki a pontos időt. A piacon elérhető nagy számú vevő közül azonban nem könnyű választani, több paramétert is figyelembe kell venni. Az egyik ilyen paraméter természetesen az eszköz ára, azonban e mellett ugyan olyan fontos az eszköz teljesítménye, pontossága, üzembiztonság és megbízhatósága is. Annak érdekében, hogy a különböző szóba jöhető vevőket ki lehessen értékelni és össze lehessen hasonlítani, kidolgoztam egy egyedi mérőrendszert. A rendszer egyedi hardware és szoftver elemekből áll, melyek segítségével hosszú távú mérések során meghatározhatóak az egyes vevők karakterisztikái.

A mérőrendszerrel több mérést is végeztem. Egyrészt összehasonlítottam azonos típusú vevőket (L86), melyek mérése során érdekes volt azt látni, hogy teljesítményük jelentősen eltér, annak ellenére, hogy még ráadásul ugyanabból a batch-ből is származtak. Másrészt terveztem egy BeagleBone SBC alapú mesterórát egy más fajta (u-blox) GNSS vevővel és ezzel is mérést, illetve összehasonlító mérést végeztem. Igen érdekes volt látni, hogy habár az új és sokkal drágább vevő talán egy kicsit jobb üzembiztonsággal működik, de időzíteni szempontból ugyan olyan, mint az először tesztelt olcsóbb társai közül a felső (jobb) harmad. A kidolgozott időzítés mérés pontosságára jellemző érdekesség, hogy a rendszerben használt mikrovezérlő kristályok frekvenciáinak hőmérséklet függését és vándorlását is szépen látni, mérni lehet.

Referenciák:

- Ádám HOLLÓS, Tamás KOVÁCSHÁZY: *Measurement System for the Performance Assessment of GNSS Receivers*. 2019 20th International Carpathian Control Conference (ICCC), Kraków-Wieliczka, Poland, 2019.
- Á. HOLLÓS: *Mesteróra Beaglebone Cape Fejlesztése*. Szakdolgozat, Budapest, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2018.

Környezeti monitorozó rendszer levegőminőség vizsgálatára

Hallgató: Kanál Attila Károly

Mentor: Dr. Kovácsházy Tamás

A levegő minősége erősen befolyásolja az emberek viselkedését. Irodalmi adatok alapján bizonyított, hogy a túlságosan meleg vagy hideg hőmérséklet, magas páratartalom, zajszint és különösen a túl magas szén-dioxid szint rossz hatással van a tanulóképességre. A referenciámként megadott Brescia-i egyetemen végzett kutatásban egy beltéri levegő minőségét vizsgáló rendszer mérési adatait elemezték, és állapították meg hogy a mérések 22.15%-ban a szén-dioxid szintje meghaladta a friss levegőnek mondható határt. A kutatásom célja egy tantermekbe telepíthető, vezeték nélküli környezeti monitorozó rendszer elkészítése volt, amivel figyelemmel kísérhető a környezeti paraméterek változása, és határértékeik túllépése. Egy ilyen rendszer segítségével megfelelő lépések tehetők a levegő minőségének javítása érdekében. A megvalósításhoz Silicon Labs Thunderboard Sense szenzorkártyák álltak a rendelkezésemre, amik rendelkeznek a mérésekhez szükséges szenzoráramkörökkel.

A megvalósult prototípus rendszer, egy Bluetooth Low Energy kommunikációs protokollon alapuló vezeték nélküli szenzorhálózat, ami alkalmas hőmérséklet, relatív páratartalom, szén-dioxid szint, TVOC (Total Volatile Organic Compound) és zajszint megfigyelésére. Az adatok gyűjtésére, megjelenítésére és feldolgozására a project előző szakaszában WEB-es alapú megoldást használtunk, ám az adatok tárolásának és kezelésének könnyítésére SNMP-ra (Simple Network Management Protocol) térünk át a CACTI hálózat-monitorozó eszköz segítségével.

A prototípus rendszer működését egy tanszéki laboratóriumban teszteltem. Két akkumulátorról működő szenzoregységet helyeztem el a szoba két végében az egész napos mérési folyamat során. Az adatgyűjtés sikeresnek bizonyult, így a kutatási időszak végére a projekt munka következő lépése a rendszer működésének a finomítása, mint például az adatgyűjtés során létrejövő LOG fájlok kezelése, extra mozgásérzékelő szenzorok implementálása és hallgatói visszacsatolás gyűjtése kérdőívek segítségével.

Referenciák

- Stefano RINALDI, Alessandra FLAMMINI, Lavinia C. TAGLIABUE, Angelo L. C. CIRIBINI: *On the Use of IoT Sensors for Indoor Conditions Assessment and Tuning of Occupancy Rates Models*. 2018 Workshop on Metrology for Industry 4.0 and IoT, Brescia, Italy, 2018.
- Attila KANÁL, Tamás KOVÁCSHÁZY: *IoT Solution for Assessing the Indoor Air Quality of Educational Facilities*. 2019 20th International Carpathian Control Conference (ICCC), Kraków-Wieliczka, Poland, 2019.

Adatvesztési modellek vizsgálata idő- és frekvenciatartományban

Hallgató: Palkó András

Mentor: Dr. Sujbert László

Az adatvesztési problémát vizsgáltuk, mely egy általánosan előforduló és elkerülhetetlen jelenség elosztott információs rendszerekben, mint például az RT-CPS alkalmazási környezetben. Az adatvesztés általánosan jelent hiányzó adatokat, érvénytelen adatokat és szinkronizációs problémákat. Az első kettő egy lehetséges matematikai leírása az indikátorfüggvény segítségével adható meg, amelyek vizsgálatát az előző munkaszakaszban kezdtük meg, és az elért eredményeket a 20th International Carpathian Control Conference-en ismertettük mentorommal, valamint azokat TDK konferencián is előadtam.

Ebben a problémakörben a mostani munkaszakaszban folytattuk a Gilbert-Elliott-modellcsalád vizsgálatát. Az előző szakaszban kidolgozott identifikációs eljárást verifikáltuk. A verifikációhoz szimulációs környezetet készítettünk, melyben az identifikációs eljárás paraméterbecslési hibájának várható értékét és szórását mértük. Vizsgáltuk ezt a hibát a modellparaméterek, a DFT-hossz és a rendelkezésre álló DFT-blokkok száma függvényében. Azt tapasztaltuk, hogy az eljárás használható, ám nem túlságosan pontos, ezért más módon közelítettük meg a problémát.

Levezettük a Gilbert-Elliott-modell autokorrelációját, és kidolgoztunk ez alapján egy várhatóan pontosabb identifikációs eljárást. Ezt az eljárást is verifikáltuk a korábbiakban ismertetett szimulációs környezetben, mellyel az eljárás nagyobb precizitását tudtuk igazolni. Az eredményeket a 2020-as I2MTC konferencián tervezzük bemutatni.

Referenciák

- András PALKÓ, László SUJBERT: *Time and Frequency Domain Description of Gilbert-Elliott Data Loss Models*. 2019 20th International Carpathian Control Conference (ICCC), Kraków-Wieliczka, Poland, 2019.
- SUJBERT, László and György OROSZ: *FFT-based spectrum analysis in the case of data loss*. IEEE Transactions on Instrumentation and Measurement 65, no. 5 (2016): 968-976.

Felhő alapú hálózati megoldások

A felhő alapú hálózatokban nemcsak a számítási erőforrások használhatóak megosztottan, hanem maga a hálózat is. A hálózati funkciók virtualizálása (Network Function Virtualization) új infrastruktúrát eredményez, emellett új alkalmazási területek kerülnek előtérbe. Ebbe illeszkedik egy jelenleg is folyamatosan fejlődő tématerület az úgynevezett szerver nélküli számítások, ahol virtuális infrastruktúra kiépítése és menedzselése nélkül vehetők igénybe funkcióhívások, programkód függvények szolgáltatásként (Function as a Service – FaaS). Kihívást jelent, hogy hogyan építhető fel ilyen funkcióhívás szolgáltatások segítségével egy komplex hálózati szolgáltatás, mint például a távvezérlés. Ezen lehetőségek feltérképezése és vizsgálata fontos kitűzött kutatási irány a különös tekintettel a megoldások teljesítményére, optimalizálási lehetőségeire és a minőségre vonatkozó garanciákra.

Cloud Networking Technologies

Cloud networking offers advanced networking services in addition to shared computing resources for cloud service customers. Using Network Function Virtualization (NFV) results in a new infrastructure, moreover, new application areas are emerging. Within this field, a dynamically expanding area is serverless computing, providing Function as a Service (FaaS). With FaaS function calls can be instantiated on demand without the burden of building and managing a virtual infrastructure. It is a challenge how complex networking services can be built with FaaS, such as remote controlling. Therefore, the investigation and evaluation of these issues is an important research goal especially for the performance profile, optimization options and for the quality guarantees of the available solutions.



Dr. Maliosz Markosz

Távközlési és Médiainformatikai Tanszék

Department of Telecommunication and Media Informatics

Maliosz@tmit.bme.hu

Serverless megoldások teljesítményvizsgálata

Hallgatók: Hegyi Gellért, Torma Kristóf

Mentor: Dr. Maliosz Markosz

A FaaS megoldások a szolgáltatást igénybe vevő ügyfelek számára több előnyt is nyújtanak: nem kell a háttér infrastruktúra kialakításával és fenntartásával foglalkozniuk, illetve a szolgáltatás számlázása az igénybevett függvény hívás ideje alapján történik, ami jellemzően nagyon rövid időtartam. Ennek következtében, mivel nincs dedikált erőforrás allokálva, a kiszolgálás teljesítménye ingadozó lehet, ezért első célkitűzésünk az elérhető nyilvános megoldások jellemzése, teljesítményelemzése, hogy meghatározzuk, mennyire alkalmasak minőségi garanciákat igénylő hálózati szolgáltatások kialakítására. A feladatok egyrészt az AWS nyilvános szolgáltatásaira, másrészt a nyílt Knative serverless keretrendszer vizsgálatára terjednek ki.

Egy komplex alkalmazás nem csak FaaS függvényekre épül, hanem kombinálva működik több különböző technológiával, úgymint pl. adatbázisok, üzenetsorok. Az adatbázisokkal való együttműködés vizsgálata során egy NoSQL adatbázist, az AWS DynamoDB-t és egy relációs PostgreSQL adatbázist kapcsoltunk egy serverless példaalkalmazáshoz. Az eredmények alapján teljesítményben a DynamoDB összességében jobban szerepelt, tehát, ha a felhasználás és az adatmodell engedi a DynamoDB egy jobb és költséghatékonyabb választás, mint egy relációs. [1]

A FaaS architektúra egy ideális felhasználás üzenetsorok (queue) menedzseléséhez, mivel ezek a használati esetek általában nagyon változó, dinamikus rendelkezésre állást igényelnek. Az AWS Simple Queue Service (SQS) egy menedzselte queue szolgáltatás, ami támogatott AWS Lambda oldalról. Ez azt jelenti, hogy képesek vagyunk egy Lambda függvénnyel egy SQS queue-t elemekkel feltölteni. Ezután egy másik Lambda függvényt rákötöttünk a queue másik végére, így amikor új elem került a queue-ba, az SQS automatikusan meghívta a párhuzamos feldolgozást is menedzselve, így a valóshoz közeli módon tudtunk több ezres elemszámmal forgalmat generálni.

A Knative keretrendszer egy Kubernetes-re épülő platform, amely serverless architektúrát biztosít. A Knative keretrendszer új platform a számos nyílt forráskódú Kubernetes-re épülő serverless platformok között, ezért fontos cél a képességeinek vizsgálata. A vizsgálatok elsősorban a függvényekhez rendelt erőforrások automatikus skálázási lehetőségeire és annak működésére vonatkoznak.

Referencia

- [1] Gellért István HEGYI, Markosz MALIOSZ, Csaba SIMON: *Measurements based performance evaluation of Function-as-a-Service Platforms*. International Conference on Big data, IoT, and Cloud Computing (ICBICC 2019), 29-31 July 2019, Rome, Italy.

Function as a Service alapú távvezérlés

Hallgató: Zsámár Olivér

Mentor: Dr. Simon Csaba

Egy komplex szolgáltatás FaaS alapú megvalósításához szükség van a függvények összekapcsolására, interakciójára. Ehhez megvizsgáltam, hogy nyílt forráskódú FaaS rendszerek függvényei milyen modellek szerint tudnak együttműködni. Az együttműködés tesztelése három modell esetén sikeres volt. Emellett több FaaS rendszer esetében elemeztem, hogy hogyan támogatja az adott rendszer a függvények közötti kompozíciót, vagyis az FaaS Workflow szervezést. Ehhez készítettem egy összefoglaló dokumentációt a rendelkezésre álló funkciókról, architektúráról és a használható külső eszközökről. A tesztelt FaaS Workflow implementációk: az OpenFaaS-hoz készített faas-flow és a Fission-höz készített Fission-Workflows voltak. Ezt a két rendszert teszteltem, ezáltal fény derült arra, hogy mely komponensek működnek, és hogy mik a hiányosságok a dokumentációhoz képest.

A komplex mintaszolgáltatás, ami kidolgozásra fog kerülni, egy drón programozott távvezérlése. Ehhez egy olyan függvénysorozaton kezdtem el dolgozni, amely egy drónból érkező információkat dolgoz fel, majd ezek alapján utána irányítja azt. A beérkező adat elég sokféle információt jelenthet (pl. video folyam, szenzorinformációk, stb.). A folyam képeiből fel kell ismerni, ha a képen egy meghatározott négyzet alakú jelölő kód (ArUco) látható, ez alapján meg kell tudnunk, hogy hova kell küldeni a drónt, és végül vezérlő parancsokat kiadni. Ehhez megterveztem egy architektúrális modellt és kerestem szabadon hozzáférhető függvénykönyvtárakat egy-egy funkció megvalósítására. Jelenleg Python API-k állnak rendelkezésre GPS helymeghatározásra és ArUco kód feldolgozására. Emellett a drón operációs rendszerének megismerése is fontos volt. A drón saját wifi hálózatot hoz létre, ebben sok porton adatokat küld és fogad. Megismerkedtem a drón működését szimuláló Gazebo 9 szimulátort, amihez szükséges több kiegészítő program is, ezek közül az egyik fontos a MavROS, amely a szimulációban futtatott programot képes MavLink üzenetökké alakítani, ennek segítségével a program a valós drónon is futhat. [3] A navigáció és a paraméterezés egyszerűsítése miatt QGroundControl szoftver használata ajánlott.

Referenciák

- [1] Swarvanu SENGUPTA: *FaaS-flow – Function Composition for OpenFaaS*. <https://github.com/s8sg/faas-flow>
- [2] Fission.io: *Fission Workflows*. <https://github.com/fission/fission-workflows>
- [3] Dronecode Project, Inc.: *PX4 Development Guide: Simulation*. <https://dev.px4.io/v1.9.0/en/simulation/>

Nyílt forráskódú Function as a Service platformok optimalizálása

Hallgató: Balla Dávid

Mentor: Dr. Maliosz Markosz

Az FaaS környezetek számos programozási nyelvet (pl. Python, Node.js, Go) támogatnak, azonban a különböző nyelvek különböző futási mintákat mutatnak. A nyílt forráskódú megoldások közül az OpenFaaS, a Kubeless, a Nuclio és a Fission is támogatja a fent felsorolt nyelvek mindegyikét. Az FaaS alkalmazásával egy alkalmazás különböző nyelveken megírt modulokból épülhet fel, ehhez azonban szükség van egy általános interfészre, amire a nyílt forráskódú platformok HTTP szerver alapú megoldást használnak, amely a felhasználó által definiált függvényeknek továbbítja a hívásokat. Python esetén az egyik legelterjedtebb megoldás a Flask HTTP szerver, amely képes a kéréseket sorosan vagy párhuzamosan feldolgozni. A Python nyelv a párhuzamos működést szálakkal vagy leszármaztatott folyamatokkal oldja meg. Fontos, hogy a szálak egy közös interpretert használnak, mely szűk keresztmetszet lehet a futás során. Leszármaztatott folyamatok esetén minden folyamat egy saját interpretert példányosít, így a fenti szűk keresztmetszetet jelentő probléma nem figyelhető meg. A Node.js által definiált HTTP szerver is lehetővé teszi a bejövő kérések soros és párhuzamos feldolgozását, utóbbi esetben a Node.js HTTP szerver előre definiált számú folyamatot indít. A Go nyelv lehetővé teszi erősen párhuzamos alkalmazások megvalósítását, ehhez a Go nyelv ún. rutinokat használ, melyek a Go keretrendszer által példányosított virtuális feldolgozó egységeken futnak.

Három jellemző függvény típust vizsgáltam: kérés-válasz (echo), számítás-intenzív és adat-intenzív függvények. Python és Node.js esetén mérésekkel igazoltam, hogy a leszármaztatott folyamatok megoldás nyújtja a legjobb teljesítményt számítás- és adatintenzív esetekben, azonban az echo függvénytípusnál a soros feldolgozás adta a legjobb teljesítményt. A méréseim alapján több soros feldolgozást biztosító függvényfuttató-környezet több nagyságrenddel is jobban teljesíthet, mint a leszármaztatott folyamatokon alapuló megoldás. Bemutattam azt is, hogy míg számításintenzív esetben a kérések soros feldolgozásával érhetünk el nagyobb teljesítményt, addig adatintenzív esetben a kérések szálakon alapuló párhuzamos feldolgozása mutat jobb eredményeket. A Go környezet esetén megmutattam, hogy elosztott esetben, a Go keretrendszer által definiált feldolgozási egységgel alacsonyabb válaszidők érhetők el, mint egy több feldolgozási egységet biztosító függvényfuttató-környezettel. [1]

Referencia

- [1] Dávid BALLA, Markosz MALIOSZ, Csaba SIMON, Dániel GÉHBERGER: *Tuning Runtimes in Open Source FaaS*. The 9th IEEE International Symposium on Cloud and Service Computing, (IEEE SC2 2019), Nov 18-21 2019, Kaohsiung, Taiwan.

Számítási felhő és minősége

A felhő-alapú informatikai megoldások a tudományos kihívások sokaságát vetik fel, többek között a rendszerek minőségével, teljesítőképességi elemzésével, tervezésével kapcsolatos kérdéseket is, melyek megválaszolása koránt sem egyszerű feladat. A téma terület kutatása számos kihívást tartogat, és több, az elosztott rendszerek területén rég ismert problémát vet fel. Így például a KPI (Key Performance Indicator) paraméterek mérését és szolgáltatási szintjének SLA (Service Level Agreement) menedzsmentjét. A kutatás célja azoknak a módszereknek, eljárásoknak a kidolgozása, amelyek a felhő alapú rendszerekben alkalmasak a KPI paraméterek mérésére, valamint a felhasználókkal kötött szerződések ellenőrzésére, menedzselésére.

A harmadik munkaszakasz legfontosabb részeredménye, hogy a BME-n kifejlesztett CIRCLE felhőmenedzsment rendszerhez egy új ütemezési algoritmust fejlesztettünk ki, ami képes figyelembe venni azokat az adatgyűjtési és mérési pontatlanságokat, amelyek a korlátozott mintavételi gyakoriságból adódnak.

Cloud Computing and it's Quality

Cloud-based IT solutions raise a multitude of academic challenges, including issues related to the quality, performance analysis and design of systems. Answering these questions, not an easy task. The research of this subject has many challenges and raises several problems that have long been known in distributed systems. For example, measuring the KPIs (Key Performance Indicators) and managing the SLAs (Service Level Agreements). The research aims to develop methods and procedures that are capable of measuring KPI parameters in cloud-based systems and controlling and managing the SLAs.

The most important result of the third project phase is the development of a new scheduling algorithm for the CIRCLE cloud management system developed at BME, which is able to take into account data collection and measurement inaccuracies due to the limited sampling frequency. The new algorithm has been implemented and successfully tested in a current version of faculty cloud (www.smallville.cloud.bme.hu).



Dr. Szeberényi Imre

Irányítástechnika és Informatika Tanszék
Department of Control Engineering and Informatics
szeberenyi@iit.bme.hu

Ütemezési algoritmusok vizsgálata

Hallgató: Belákovics Ádám

Mentor: Dr. Szeberényi Imre

A részfeladat célja a felhő menedzsment rendszerekben alkalmazott ütemezési algoritmusok hatásának vizsgálata a KPI paraméterekre. A feladat a gyakorlatban közvetlenül megjelenő probléma tudományos igényű megoldásához kapcsolódik. A konkrét probléma a BME-n kifejlesztett felhő menedzser (CIRCLE) kapcsán fogalmazódott meg. A rendszer terhelésének növekedésével egyre komolyabb kérdéssé vált az objektív módon mérhető hatékonyság és szolgáltatás minősége. A felhő rendszerek működésének egyik szokásosan kiemelt mérőszáma a virtuális gépek (VM) indítási ideje, ami alapvetően meghatározza a rendszer dinamikus terhelhetőségét. A VM indítási időt legnagyobb mértékben az ütemezési algoritmus befolyásolja, ezért a részfeladat keretében az ütemező algoritmusra fókuszáltunk. A konkrét infrastruktúra használati eseteiből kiválasztottuk azokat az eseteket, melyekben a jelenlegi algoritmus rosszul teljesít. A naplók elemzésével sikerült azokat a paramétereket megtalálni, amelyek kevésbé hatékony irányba vitték az algoritmust. Ezt követően különböző ütemezési algoritmusokat teszteltünk szimulációs környezetben, majd CIRCLE rendszerben megvalósítottuk a legjobbnak tűnő megoldást. Alapvető cél az volt, hogy kifejlesszünk egy új ütemező algoritmust, ami a jelenleginél jobb eredményt ér el a virtuális és fizikai gépek egymáshoz rendelése során. Megállapítottuk, hogy különösen fontos az ütemező algoritmus bemeneti paramétereinek pontossága, naprakészsége. Elavult információk alapján ugyanis a legjobb algoritmusok is rossz döntéseket fognak hozni. Ugyanakkor mért mennyiségek (terhelés, ram és diszk allokáció) túl gyakori mintavétele viszont jelentős terhelésnövekedést okozhat, különösen nagyméretű elosztott rendszerben.

Végül egy olyan prediktív algoritmust fejlesztettünk ki, ami nem csupán a bemeneti paraméterek alapján hoz döntéseket, hanem a korábbi döntéseket is figyelembe veszi. Elért eredményeinket egy tudományos konferencián [1], valamint egy szakdolgozat keretében publikáltuk [2].

Referenciák

- [1] Ádám BELÁKOVICS, Arnold CZÉMÁN, Imre SZEBERÉNYI: *Designing and testing vm allocation algorithms for the circle cloud manager*. Proceedings of the (CS)² – The 11th Conference of PhD Students in Computer Science, 2018. június.
- [2] Ádám BELÁKOVICS: *A CIRCLE felhőmenedzser továbbfejlesztése*. Szakdolgozat, BME VIK 2018.

Szoftverrádió-alapú IoT fejlesztőrendszer

A dolgok internete (Internet of Things, IoT) napjainkban rohamosan terjed. Az IoT az alkalmazások jelentős részében valamilyen hagyományos vezeték nélküli adatátviteli eljárásra épül. Ennek a kutatásnak a célkitűzése egy tudományos és demonstrációs célú, szoftverrádiós (Software Defined Radio, SDR) alapon felépülő, költséghatékony és skálázható platform kidolgozása, és SDR keretrendszerhez illesztése. A platform lehetővé teszi, hogy a jelenleg elterjedt IoT vezeték nélküli átviteli eljárásoknál hatékonyabban lehessen nyílt technológiákkal, valamint újszerű kiberbiztonsági, modulációs, kódolási és közeghozzáférési eljárásokkal lehessen kísérletezni, új módszereket implementálni, és a javasolt módszerek teljesítőképességét a gyakorlatban mérésekkel vizsgálni. A kutatás részét képezi egyes speciális alkalmazások vizsgálata, különös tekintettel az ún. masszív többantennás (massive MIMO) technológiák alkalmazására, kutatására.

SDR-Based IoT Prototyping Platform

Internet of Things (IoT) applications are getting ubiquitous in our lives. Such systems typically rely on some traditional wireless communication technology. The present research aims at creating a comparatively cost-effective, scalable, Software Defined Radio (SDR)-based research and demonstration platform, suitable for integration into well-known SDR development frameworks. The proposed platform enables us to study wireless communication schemes that are consist of open technologies and are more efficient than today's industrial standards. We will be able to focus on physical layer security research, and improvements of modulation, error-control coding, and multiple access schemes. The platform enables experimentation, implementation, and measurement-based real-world performance evaluation of the proposed schemes. Currently, we focus on the research of specific applications, with emphasis on massive multi-antenna (MIMO) research.



Dr. Horváth Péter

Szélessávú Hírközlés és Villamosságtan Tanszék
Department of Broadband Infocommunications
and Electromagnetic Theory
hp@hvt.bme.hu

Masszív MIMO feldolgozás IoT alkalmazásokban

Hallgatók: Csathó Botond Tamás, Szabó Lóránt Zsolt

Mentor: Horváth Bálint Péter

A vezetékek nélküli infrastrukturális hálózatok kapacitásnövelésében az utóbbi néhány év újabb áttörést hozott az ún. masszív többantennás (massive MIMO) technikák kidolgozása és alapos kutatása. A masszív MIMO megközelítés a hálózat bázisállomásaiban, illetve átjáróiban nagy számú, akár több tíz vagy több száz egyedi antenna és a hozzájuk tartozó rádiófrekvenciás feldolgozó lánc felhasználásával éri el, hogy a hagyományos többfelhasználós MIMO technikáknál robusztusabb, nagy mértékben skálázható hozzáférést biztosíthassunk nagy számú, jellemzően kis aktivitású felhasználónak. A vonatkozó kutatások elsősorban a cellás mobilhálózatokra fókuszálnak, azonban a megközelítés az IoT-világban is eredményesen hasznosítható.

A masszív MIMO kutatás egyik újszerű, ígéretes iránya az ún. nagy apertúrájú masszív MIMO rendszerek kutatása, amelyben a bázisállomás antennái egymástól távol helyezkednek el, például egy irodaház ablakaiba helyezik el azokat. Ilyen esetben az antennarendszer sugárzó közeletere több száz méterre vagy akár kilométerekre is nyúlhat. A közel térben a hullámterjedés sajátosságai jelentősen eltérnek a hagyományos távöltéri esettől. Emiatt a módszerek teljesítményvizsgálatahoz elengedhetetlen hagyományos rádiócsatorna-modellek sem alkalmazhatók közvetlenül. Az alprojekt keretében a nagy apertúrájú rendszerekben alkalmazható csatornamodellezési eljárásokat és a nagy apertúrájú rendszerekkel elérhető teljesítményt vizsgáljuk. A síkhullám-közelítéseken alapuló modellezési módszerekkel szemben pontos elektromágneses térszámítási módszereket (pl. síkhullám-felbontást) használunk, így jellemezve az elektromágneses tér és a környezet kölcsönhatásainak statisztikáit.

Az alprojekt keretében tanulmányoztuk a masszív MIMO rendszerek egyik kritikus jelfeldolgozási lépését, a rádiócsatorna állapotának a becslését is. Alapvető fontosságú, hogy a csatornabecslés kis varianciával, kis számítási komplexitással elvégezhető legyen, pl. programozható logikai (FPGA) áramkörök segítségével. A kutatás során magas szintű logikai szintézis alkalmazásával vizsgáltuk a csatornabecslés implementációs kérdéseit.

Referencia

- [1] CSATHÓ Botond Tamás: *Felmenő ági jelfeldolgozás masszív MIMO rendszerekben*. OTDK Informatika Szekció (2. díj), 2019.

Időkritikus 5G architektúrák

A felhasználók jelenlegi és várható jövőbeli igényeinek kiszolgálására megoldást jelenthet a kisméretű cellák alkalmazása, célszerűen Radio over Fibre (RoF) kiépítésével, mely technológia a hálózat késleltetését is minimalizálja. Az alkalmazott modulációs megoldások, a felhasznált optikai komponensek, a hullámhossz menedzsment, a topológia kialakítása, mind azt a célt szolgálja, hogy a CPRI (common public radio interface) által kínált megoldást túlszárnyaljuk.

A hallgatói feladat fő célkitűzése az 5G rendszerek fejlesztésében nagy szerepet játszó új technológiák felhasználásával a hálózat architektúrájának optimalizálása. Mindezt összetett módon, modellek alkotásával és a szimulációs eredmények mérésekkel történő igazolásával. A harmadik mérőföldkő során az 5G hálózatokban az optikai szegmens kiterjesztésére, a szabadtéri szakaszon való optikai megközelítés vizsgálatára fektettük a hangsúlyt. Ezen kívül kiemelten foglalkoztunk a kommunikáció biztonságosságának kérdéseivel, ehhez kvantumos elvű véletlenszám generátor elméleti és gyakorlati vizsgálatát végeztük el. A feladatokhoz szerteágazó előzetes ismeretek szükségesek, beleértve az összetett optikai-mikrohullámú rendszereket és optikai mérés technikát.

5G time critical architectures

For the evolution toward 5G, it is envisioned that optical networking will play a major role in supporting the requirements while reducing the deployment costs through the introduction of novel converged fiber-wireless networks. Radio over Fiber (RoF) technology will support the applied small cells and also minimizes network delays. The applied modulation solutions, the selected optical components, the wavelength management and the topology of the network are all designed to overcome the solution offered by the common public radio interface (CPRI).

The main goal of the student task is to investigate network optimization in time-critical systems. It is a complex work, scientific models will be evolved, and the simulation results will be validated by experimental work. In the third year, the students focused on the optical-wireless technologies. On the other hand, the students investigated the security and reliability of the network. They theoretically and experimentally investigated a quantum random number generator. The realized measurement equipment gives the opportunity for the validation of the previous simulation results. The student tasks require extensive prior knowledge, including complex optical-microwave systems and radio propagation models.



Gerhátne Dr. Udvarý Eszter

Szélessávú Hírközlés és Villamosságtan Tanszék
Department of Broadband Infocommunication
and Electromagnetic Theory
udvary@hvt.bme.hu

Járművek közti kommunikáció látható fény segítségével

Hallgató: Kovács Zoltán Márk

Mentor: Gerhátné Dr. Udvary Eszter

A hallgatói kutatási feladat célja egy olyan látható fényű alapokra épülő kommunikációs platformot létrehozása, amely alkalmas az 5G hálózatokban alternatív módon alkalmazható VLC (Visible Light Communication) technológia tesztelésére. A minta rendszer két 1:24 méretarányú autómoddell közti kommunikációt valósít meg. A feladat részét képezi a mechanikai konstrukció kialakítása, a kommunikációs egység és vezérlőegységek megtervezése, programozása, valamint az elkészült áramkörök rögzítése az autómódelleken belül, és a tápellátás kialakítása mindkét áramkör részére. A távolság mérését egy lézeres távolságmérő szenzor végzi, amelyet nagy pontosság jellemez, illetve kisebb térszögben detektálja a céltárgyat, mint az ultrahangos távolságmérő. A platform szoftveres támogatását Arduino NANO biztosítja, amely olcsó, egyszerűen programozható, kisméretű fejlesztőplatform, és rendelkezik minden szükséges funkcióval, mint a PWM jel létrehozása, UART csomagok küldése, fogadása, programozható digitális/analóg pinek. A munka eddigi eredményeként elkészült két adó- és vezérlőáramkör, amely látható fényt használva képes UART csomagok küldésére, és DC motor meghajtására. További funkciói az akkumulátorok feszültségének visszaolvasása, DC motor áramfelvételének mérése. Emellett elkészült két vevőáramkör, amely képes a látható fénnel érkező UART csomagok fogadására, majd átalakítja az Arduino által feldolgozhatóvá.

Referenciák

- [1] K. SIDDIQI, A. D. RAZA and S. SHEIKH MUHAMMAD: *Visible light communication for V2V intelligent transport system*. 2016 International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom), Graz, Austria, 14-16 Sept. 2016.
- [2] Dahmani MOHAMMED, Dounia Khawther Dihya BOURZIG, Meche ABDELKIM, Keche MOKHTAR: *Digital data transmission via Visible Light Communication (VLC): Application to vehicle to vehicle communication*. 2016 4th International Conference on Control Engineering & Information Technology (CEIT), Hammamet, Tunisia, 16-18 Dec. 2016.
- [3] KOVÁCS Zoltán Márk: *Járművek közti kommunikáció látható fény segítségével*. TDK dolgozat, 2019.

Valós idejű bitgenerálás kvantum-véletlenszámgenerátorral

Hallgató: Marosits Ádám

Mentor: Schranz Ágoston

A hallgatói kutatási feladat célja egy kísérleti stádiumban lévő, optikai elven működő kvantum véletlenszám-generátor (QRNG) jeleinek feldolgozása és a fizikai jelek alapján generált véletlenszámok megjelenítése. Az 5G hálózatokban is kiemelt fontosságú megbízhatóság és biztonságos kommunikáció miatt szükséges a jó minőségű véletlenszámok előállítása. A feladatban szereplő QRNG egy erősített spontán emisszió alapuló megoldás, ahol a kvantumos optikai jelenség által keltett zaj biztosítja a generált számsorozatot. Az optikai-elektromos átalakító kimenetén megjelenő nyers elektromos jel utófeldolgozást igényel, majd a generált számsorozaton véletlenség tesztet kell futtatni és végül a megfelelő minőségű bitsorozatot tárolni kell, illetve szükség esetén meg is kell jeleníteni.

Az első verzió során egy C nyelven felprogramozott ESP8266 alkalmazásával egy WiFi hálózaton keresztül elérhető weboldal készült el. A mikrokontrollerre kapcsolódva a rendszer képes megadott számú véletlen bitet generálni, melyet egy weboldalon megjelenít. A megoldás lassú, és korlátozott elérésű. A továbbfejlesztett változatban egy ESP8266 mintavételezi az optikai-elektromos átalakító kimeneti jelét, és UART-on keresztül kommunikál a Raspberry Pi 3-al. Egy PHP program létrehozta a weboldalt, amely kommunikál az SQL adatbázissal, illetve lefuttat egy Python kódot is. Az ESP8266 programja a soros port aktívra válása után beolvassa a kért bitek számát, feltölt egy tömböt az analóg bemenetről, az ezekből képzett átlaghoz komparálva feltölti a tárolónkat bitekkel. Ekkor kiválaszt a kért bitnek megfelelő mennyiségű adatot, majd soros porton továbbítja a Raspberry Pi felé. A Raspberry programozása során HTML, CSS, PHP, Python és SQL nyelv használatára is szükség volt.

A jelenlegi rendszer több irányban is tovább fejleszthető. Egyrészt egy saját rendszerhez optimalizált ADC-vel rendelkező áramkör jobb minőségű véletlen biteket képes generálni. Másrészt a tárolási módszerek javításával növelhető a weblapon megjelenített bitek száma. Harmadrészt a jelenleg csak helyi hálózaton működő megoldást ki kell terjeszteni. Végül az utófeldolgozás javítása jobb minőségű véletlen bitek létrehozására ad lehetőséget. Az elkövetkezendő időszak feladata ezeknek a megvalósítása.

Referenciák

- [1] Á. SCHRANZ, Á. MAROSITS, E. UDVARY: *Effects of Sampling Rate on Amplified Spontaneous Emission Based Single-Bit Quantum Random Number Generation*. 21st International Conference on Transparent Optical Networks (ICTON), DOI: 10.1109/ICTON.2019.8840188.
- [2] MAROSITS Ádám: *Erősített spontán emisszió alapuló kvantum véletlenszám-generátor*. TDK dolgozat, 2019.

Neurális hálózatok alkalmazása fading jelenségek automatikus detekciójában

Hallgató: Márton Boldizsár

Mentor: Dr. Csurgai-Horváth László

A 2013 óta a Föld körül keringő Alphasat műholdat az Európai Űrügynökség az Inmarsattal együttműködve állította pályára, amely két frekvencián, 19,701 GHz-en és 39,402 GHz-en sugároz modulálatlan szinuszos jelet, amelyet a BME-HVT-n található vevőállomás folyamatosan rögzít. A vett jel szintje az időjárási viszonyok függvénye és a pontos fading esemény azonosítás kritikus a későbbi adatelemzés szempontjából. A kutatás célja ezen események felismerése mesterséges intelligencia alkalmazásával. A másodpercenként mintavételezett jelszint értékek alapján neurális hálózatok segítségével kívánunk különbséget tenni esőesemények és nem esőesemények között. A feladatmegoldást előkészíti a neurális hálózatokkal való megismerkedés, amelyet egy tényleges hálózat betanítása és valós adatokon történő tesztelése követ. A betanításhoz szükséges adatokat a BME-HVT vevőállomása által rögzített idősorok szolgáltatják. Az elsőként elért eredmények a H-SPACE 2019 konferencián [1] kerültek publikálásra. Folytatásképpen kutatás történt annak felderítésére, miként érdemes megválasztani és létrehozni a nyers adatokból a hálózat számára hasznos betanító adatokat.

A neurális hálózat mérési adatokkal történő betanítása utáni feladat egy független adatsoron történő tesztelés volt. A neurális háló segítségével megkereshetők az adatsorban azok az eseményeket, amelyeket valamilyen atmoszférikus hatás okoz. Ezt egy Matlab script segítségével, változtatható szélességű ablaknak a mért adatsoron történő végigolásával lehet megvalósítani, felhasználva a háló által nyújtott klasszifikáció lehetőségét.

A későbbi eredmények egy angol nyelvű előadásban kerültek összefoglalásra, amely ismertette az addig elvégzett munkát és előrevetítette a további haladási irányt. Az előadást a BME-HVT által szervezett Tanszéki Szemináriumon hangzott el [2].

Referenciák

- [1] László CSURGAI-HORVÁTH, Boldizsár MÁRTON: *Satellite Fading Classification with Artificial Intelligence for 5G*. 5th International Conference on Research, Technology and Education of Space (H-SPACE 2019), 27-28 February, 2019, Budapest, Hungary, ISBN 978-963-7367-18-2.
- [2] Boldizsár MÁRTON and László CSURGAI-HORVÁTH: *Applying machine learning to recognize fading on satellite radio channel*. Mesterpróba 2019, Budapest, Hungary, 19.06.2019.

A digitális infrastruktúra szolgáltatási képességeinek fejlesztése és automatikus üzemeltetésének megalapozása

A közeljövő információs rendszereiben különféle alkalmazási környezetekben (közlekedés, ipar, monitorozó- és beavatkozó rendszerek) rendkívüli nagyszámú és nagy sűrűségben jelenlévő, részben mozgó végpont hálózati kiszolgálása válik szükséges-sé. A hagyományos hálózati infrastruktúrák erre a célra skálázhatósági, flexibilitási, megbízhatósági és késleltetési problémák miatt nem alkalmasak.

A kutatás célja az 5G rendszerek szolgáltatási képességeinek fejlesztése érdekében

- az „edge computing” architektúra vizsgálata és fejlesztése keretében algoritmusok kidolgozása a futtatandó virtuális funkciók terhelésfüggő dinamikus el- és áthelyezésére, továbbá a szolgáltatási minőségjellemzőkre gyakorolt hatások analitikus és szimulációs vizsgálata, valamint
- a hálózati erőforrások dinamikus szolgáltatáshoz rendelésének („dynamic end-to-end network slicing”) automatizálhatóságát megalapozó analízis-, tervezési és döntési modellek, algoritmusok kidolgozása, illetve
- az 5G rádiós interfész vizsgálata az ultra sűrű és gépi végpontok skálázhatóságának, rugalmasságának, robusztusságának figyelembevételével; teljesítőképességi paraméterek becslése analitikus és szimulációs modellek segítségével.

Methods to Enhance the Service Capability of Digital infrastructure

In the near future, 5G networks will be deployed for various vertical industries (automotive, energy, food and agriculture, city management, etc.). The service provisioning of 5G networks introduces many new challenges and issues like scalability, flexibility, reliability and latency.

Therefore, the aims of the pursued research include

- the performance evaluation of mobile edge computing architecture with regard to virtual network functions, the quality of services and demands from subscribers,
- the development of methodology to investigate the viability of dynamic end-to-end network slicing solutions for vertical industries,
- the investigation of 5G radio interfaces taking into account the requirement of a huge number of endpoints, the scalability and the robustness of solutions.



Dr. Do Van Tien

Hálózati Rendszerek és Szolgáltatások Tanszék
Department of Networked Systems and Services
do@hit.bme.hu

Erőforrás felhasználásának optimalizációja aktor-alapú rendszerekben

Hallgató: Nguyen Hai

Mentor: Dr Do Van Tien

Manapság egyre több digitális készülék képes valamilyen módon csatlakozni hálózathoz. Az ezen eszközök mögötti globális hálózati infrastruktúrát nevezzük a dolgok internetének (Internet of Things, IoT), mely gyűjtőfogalomként tengernyi különféle technológiát foglal magába (V2V, RFID, WSN), és amely a rengeteg különféle eszközből egy hatalmas elosztott és heterogén rendszert alkot.

A nagy elosztott rendszerek szerver oldali megvalósítására az aktor modell nyújthatja a megoldást mely az egymással üzeneteken keresztül kommunikáló aktorokra alapszik. Ez az üzenetváltásos kommunikáció és az aktorok által alkotott hierarchikus struktúra kiválóan illeszkedik az IoT rendszerek szerkezetéhez.

Az Akka egy Scala-ban íródott eszköztár, mely aktor-alapú rendszerek implementálására lett létrehozva. Az Akka-ban az aktorok között a szálak elosztásáért az ún. Dispatcher felel, mely egy szálkészletből rendel szálakat az aktorokhoz adott ütemezési stratégia alapján. A rendszer teljesítménye nagyban függ a szálkészlet méretétől, hiszen kis méret esetén az aktorok nem fogják tudni kiszolgálni a kéréseket időben, nagy méret esetén pedig fölöslegesen használhatjuk fel az erőforrásainkat, ráadásul az is megeshet, hogy a nagy mennyiségű szálak esetén a rendszer összkapacitása csökken a kontextusváltások miatt.

Munkánk során arra voltunk kíváncsiak, hogy a szálkészlet optimális méretét meg lehet-e találni a gépi tanulás módszereivel. A vizsgálatokhoz egy szimulált környezetet készítettünk, melyben adott méretű IoT forgalmat kellett kiszolgálni változtatható szálkészlet mellett. A szálak indítását és leállítását egy ágens végezte, melynek feladata az volt, hogy megtanulja az optimális döntést a rendszer adott állapotában. Az ágenset egy ún. actor-critic algoritmussal tanítottuk, és a célfüggvényt úgy állítottuk be, hogy az optimális esetben minimalizálja a szálak számát úgy, hogy az igények várakozási ideje egy adott érték alatt maradjon. Megfelelő mennyiségű tanítás után az ágensnek sikerült megtalálnia az optimális szálkészlet méretet abban az esetben, ha az adott forgalom mellett ez lehetséges volt. Azokban az esetekben, ahol a forgalom nagysága akkora volt, hogy semmilyen módon nem lehetett a várakozási időt szinten tartani, az ágens ezt az értéket megpróbálta minimalizálni.

Referencia

- Roestenburg, R., Bakker, R., & Williams, R. (2015). Akka in Action (1st ed.). Greenwich, CT, USA: Manning Publications Co.

Adatközponti hálózati topológiák hatása virtuális hálózati funkciók szolgáltatásláncainak megvalósítására

Hallgatók: Szegeczky Imre, Szilágyi Nándor

Mentorok: Jakab Tivadar, Gódor Győző

A hagyományos adatközponti hálózati topológiák a távoli felhasználó és a számítási erőforrások (szerverek) közti robusztus, hibatűrő – vertikális – kapcsolat biztosítását célozzák. A hálózati funkciók virtualizálására alapozottan, a virtuális hálózati funkciókból épített szolgáltatáslancok az adatközpontokon belüli – horizontális – szerver-szerver kapcsolatokat használják.

Munkánk során egyszerű, protokoll- és gyártmányfüggetlen modellekre alapozottan vizsgáltuk, hogy hogyan befolyásolják a különböző tipikus adathálózati topológiák (Fat Tree, DCell, BCube, stb.) a virtuális hálózati funkciókból épített szolgáltatáslancok által nyújtott szolgáltatások minőségét, rendelkezésreállását.

Elemzéseink célja olyan topológia módosítások azonosítása, melyek a minél kisebb többleterőforrás-ráfordítással, illetve számítási kapacitásvesztéssel képesek a szolgáltatáslancok jellegéhez jobban illeszkedő feltételeket biztosítani.

Véletlen hibák okozta hálózati vészhelyzeti incidensek kezelése szolgáltatói IP hálózatokban

Hallgató: Farkas Levente

Mentor: Jakab Tivadar

Egy nagyméretű és komplex IP hálózatban egy nem várt hálózati esemény – pl. kábelvágás, router vagy vezérlő eszköz szoftver-, illetve hardverhiba – olyan (lánc-) reakciókat eredményezhet, amelyek a szolgáltatások működésében degradációt vagy teljes leállást okozhatnak. Ilyen esetben szükség van a különböző rendszerek szakértőinek koordinálására a szolgáltatások stabilizálása, helyreállítása érdekében. Az ilyen kritikus hatású véletlen hálózati hibák elhárítását dokumentáló Post-mortem Reportok feldolgozása alapján megismerhetők és elemezhetők a hálózati incidensek kezelési folyamatai, melyek értékelése jelentősen hozzájárulhat a digitális infrastruktúra szolgáltatási képességeinek fejlesztéséhez és automatikus üzemeltetésének megalapozásához.

Munkánk célja egy olyan gyakorlatban alkalmazható módszertan kidolgozása és gyakorlati alkalmazása, mely alkalmas a hálózati vészhelyzeti incidenskezelés folyamatainak leírására, és ezen folyamatok összehasonlító értékelésének támogatására. A feldolgozott és értékelt incidenskezelési folyamatokból kinyerhető információk és tapasztalatok az automatikus üzemeltetés konfigurációmenedzsment és hibamenedzsment modelljeinek és módszereinek kidolgozásához járulhatnak hozzá.

Járműkommunikációs hálózatok tervezési, elemzési, szimulációs modelljei és módszerei

Hallgatók: Lévay Mátyás, Lőrinc Dániel,
Rénes Dániel, Wippelhauser András

Mentorok: Varga Norbert, Jakab Tivadar és Knapp Gábor

Az ETSI G5 ITS V2I/V2N járműkommunikációs hálózatok 3D terepmodellekre alapozott rádiós tervezésének modellek paraméterezéséhez szükséges a rádiós átviteli jellemzők fogadó oldali jelerősségre alapozott méréséhez mérési célú üzenettípust definiáltunk, implementáltunk, valamint az egyszerű, olcsó végkészüléken hozzáférhető Android V2X Packet Analyzer alkalmazást továbbfejlesztettük, kibővítettük.

A járműkommunikációt támogató hálózati szolgáltatások véletlen hibák hatását figyelembevevő jellemzésére adaptáltuk, továbbfejlesztettük és kibővítettük a hagyományos hálózatmegbízhatóság modellezésére szolgáló eszköztárat.

A kooperatív járműkommunikációban az ETSI ITS G5 járműkommunikációs hálózatokkal szemben fennálló szolgáltatásminőségi követelmények szimulációs elemzésére, a kooperatív járműkonvojok menetvezérlési algoritmusainak vizsgálatára keretrendszert fejlesztettünk (Vein, OMNeT++ nevű, C++ alapú, diszkrét idejű, eseményvezérelt szimulációs keretrendszer, a létező hálózati és kommunikációs protokollra vonatkozó modellgyűjteményeket elemezve a járműkommunikáció területén definiált célok megvalósítására legalkalmasabb modellek kiválasztásához, integrálásához). Ehhez kapcsolódik az artery (<https://github.com/riebl/artery>) szimulációs keretrendszer továbbfejlesztése is: a szimulációs keretrendszer jelenlegi fizikai modelljének lecserélése egy alapú valósághű rádiós modellre (<https://github.com/u-blox/ubx-v2x>). Az így továbbfejlesztett keretrendszer alkalmas a 802.11p protokoll valósághű szimulációjára, amivel ETSI ITS G5 rendszerek valós körülmények közti helytállóságát tudjuk alátámasztani 802.11p fizikai réteg használata esetén.

Az ETSI ITS G5 járműkommunikációs szolgáltatások működésének automatizálásához járul hozzá munkánk a városi kereszteződések sáv szintű pontosságú feltérképezésével, amelyet a járművek által szolgáltatott adatok összegyűjtésével, mesterséges intelligencia alapú feldolgozásával valósítunk meg.

Ennek első fázisa egy egyedileg kialakított, SUMO alapú forgalom szimulációra támaszkodó, különböző MI algoritmusok modul szintű cseréjére és futtatására képes fejlesztő- és tesztkörnyezet tervezése és implementálása, melyben felügyelet nélküli, félig felügyelt, és felügyelt tanulás alapokon működő trajektória klaszterezési megoldások vizsgálatára és összehasonlítására egyaránt nyílik mód.

**Intelligens adatalemzés
(C pillér)**

Intelligens adatelemzés

(C pillér)

A projekt „C” pillére, az „Intelligens adatelemzés” témában olyan alap- és alkalmazott kutatási területekkel foglalkozik, melyek közös célja, hogy a mesterséges intelligencia által nyújtott módszerekkel, új típusú megközelítéssel próbáljon az adatvezérelt rendszerek elemzési tevékenységeiben segítséget nyújtani oly módon, hogy gyorsabb, pontosabb, vagy hatékonyabb megoldással szolgáljon. A „C” pillér négy alprojektből áll, amelyekben ugyanaz a trend figyelhető meg, mint a nemzetközi szakirodalomban és az ipari alkalmazásokban, azaz a mesterséges intelligencia akadémiai jellegű kutatását és ipari fejlődését egyre inkább a mélytanulás dominálja.

Az egyik alprojekt („Gyengén koherens sokdimenziós idősorok modellezése mély tanulással”) a konvolúciós és rekurrens neurális hálózatok mellett vizsgálja többek között a Hierarchikus Temporális Memória, a Variational Autoencoder és Transformer hálózatok tulajdonságait, kibővítésének és alkalmazásainak a lehetőségeit annak érdekében, hogy az idősor modellek számításgényét csökkentse.

Másikban („Intelligens adatelemzés kiberfizikai (CPS) rendszerekben”) a neurális hálók komplexitásának automatizált redukciója a cél, ugyanis egyrészt a feladatmegoldások során nincs mindig elegendő megfelelő minőségű tanító mintakészlet, melyből a hálózatok képesek tanulni, másrészt a gépi tanulás általánosító képességére vonatkozó elvárás, illetve a számítási erőforrásigény sem mindig teljesíthető.

A természetes nyelvfeldolgozáshoz tartozó összefoglaló generálásnál, valamint az implicit visszacsatoláson alapuló kollaboratív szűrő ajánlórendszereknél – mint ahogy azt az egyik alprojektben („Nagy adathalmazok komplex struktúráinak előrejelzése az adattudomány módszereivel”) bemutatjuk – szintén szükség van mesterséges intelligencia korszerű módszereire, azaz a mély neurális hálókra, ezekkel ugyanis megvalósítható a komplex struktúrájú adathalmazokban a nemlineáris összefüggések feltárása és megtanulása.

Neurális hálózattal a PET (pozitronemissziós tomográfia) rekonstrukció is tovább fejleszthető, az egyik alprojekt („Döntéstámogató módszerek dinamikus orvosi képalkotó berendezésekhez”) keretében ugyanis azt vizsgálják, hogy mély neurális hálózatok segítségével ennek a rekonstrukciónak mely lépéseit lehet javítani. Ehhez először egy olyan zajszűrő neurális hálózatot fejlesztettek, amely az iteratív algoritmus által számított zajos képek jel/zaj arányát növeli az anatómiai régiók határainak figyelembevételével; majd a modellparaméterek gyors és pontos becslése szintén neurális hálózattal megoldható.

Intelligent data analysis (Pillar C)

The pillar “C” of the project, “Intelligent Data Analysis”, deals with basic and applied research areas that have the common goal to assist analytical activities of data-driven systems by Artificial Intelligence methods and by a new type of approach, to provide faster, more accurate or more effective solution. Pillar C consists of four subprojects, which follow the same trend as international literature and industrial applications, i.e. academic research and industrial development in artificial intelligence is increasingly dominated by deep learning.

In addition to convolutional and recurrent neural networks, one of the subprojects (“Modeling weakly coherent multidimensional time series with deep learning”) researches the properties, extensions, and applications of Hierarchical Temporal Memory, Variational Autoencoder, and Transformer networks to reduce computational time of time-series models.

At the other subproject (“Intelligent Data Analysis in Cyber-Physical Systems”), the aim is the automatic reduction of the complexity of neural networks, because on the one hand, there is not always enough high-quality training data sets from which the networks can learn; on the other hand, the expectation of machine learning generalization and computing resources may not always be met.

Summary generation, as a part of natural language processing, and collaborative filter with implicit feedback in the recommendation systems – as presented in one of the subprojects (“Predict the complex structures of large data sets with the methods of data science”) – also require advanced methods of artificial intelligence, i.e. deep neural networks, because they can be used to discover and learn nonlinear relationships in complex data sets.

The neural network can also be used to improve PET (positron emission tomography) reconstruction, as one of the subprojects (“Decision support methods for dynamic medical imaging systems”) investigates which steps of this reconstruction can be improved with the help of deep neural networks. They first developed a noise-filtering neural network that increases the signal-to-noise ratio of noisy images computed by the iterative algorithm considering the boundaries of anatomical regions; then a quick and accurate estimation of the model parameters can also be solved using a neural network.



Pillér koordinátor: Dr. Szűcs Gábor

Távközlési és Médiainformatikai Tanszék

Department of Telecommunications and Media Informatics

szucs@tmit.bme.hu

Gyengén koherens sokdimenziós idősorok modellezése mély tanulással

Számos valós életbeli folyamat sokdimenziós időbeli struktúrát követ, ahol a különböző dimenziók csak részben, sok esetben kis mértékben függenek egymástól. Ilyen adatfolyam például a pénzügyi idősorok, IoT (Internet of Things) hálózatok adatai, az EEG jelfolyamok vagy például gyártósori adatfolyamok.

A ma már klasszikusnak számító idősor elemző és modellező (pl. autoregresszív) eljárások kevés rendelkezésre álló adat esetén jellemzően jól teljesítenek, azonban az adatmennyiség növekedésével a gépi tanulás, különösképp a mély tanulás alapú eljárások nagyobb pontosságot eredményezhetnek. Ezek a módszerek jellemzően csupán implicit módon kezelik a sokdimenziós idősorok dimenziói közötti gyenge koherenciát. A kutatás célja olyan adatvezérelt módszerek vizsgálata és kidolgozása, melyek alkalmasak a jelenlegi modelleken túlmutatóan, idősorok különböző dimenziói közötti összefüggések explicit modellezésére. Az egy dimenziós konvolúciós és rekurrens neurális hálózatok mellett vizsgáljuk többek között a Hierarchikus Temporális Memória, a Variational Autoencoder és Transformer hálózatok tulajdonságait, kibővítésének és alkalmazásainak a lehetőségeit.

Modeling weakly coherent multidimensional time series with deep learning

Many real-world processes produce multidimensional temporal data, in which different dimensions are dependent on each other partly or weakly only. Such data are financial time series, data of IoT (Internet of Things) networks, EEG datafeeds or data of production line devices.

In case of a smaller dataset, traditional approaches, like autoregressive models, may be a proper choice. However, data-driven machine learning techniques, especially deep learning based approaches are likely to achieve better modeling performance in case of larger dataset, still, weak coherence in multidimensional data are considered implicitly only. The goal of this research is to elaborate and develop novel methods in terms of theoretical approach, performance, accuracy or application to be able to explicitly model weak coherence in multidimensional time series. We investigate the features and possible extensions and applications of one-dimensional convolutional and recurrent neural networks, furthermore of complex architectures, like Hierarchical Temporal Memory, Variational Autoencoders and Transformers.



Dr. Gyires-Tóth Bálint

Távközlési és Médiainformatikai Tanszék

Department of Telecommunications and Media Informatics

toth.b@tmit.bme.hu

Hierarchikus Temporális Memória alapú idősor modellek számításgényének csökkentése

Hallgató: Pilinszki-Nagy Csongor

Mentor: Dr. Gyires-Tóth Bálint

Az idősorok modellezése, predikciója fontos szerepet tölt be az ipari alkalmazásokban. A gépi tanulás számos módszert kínál idősorok tanulására, ezek közül az egyik legelterjedtebb a Long Short-Term Memory (LSTM) hálózat. Az LSTM nagy áttörést jelentett a szekvencia modellezésben, segítségével sikeresen modellezhetők hosszabb adatsorok. Az eddigi sikerek ellenére a hálózat teljesítménye más szempontokból (pl. az online, one-shot és few-shot tanulásnál) nem túl jó. A jelenlegi módszerek további hátránya a nagy számításgény, aminek oka a nagy változószám, illetve az iteratív tanulási folyamat.

Egy kevésbé elterjedt módszer a Hierarchikus Temporális Memória (HTM), amely ugyancsak egy neurális hálózat, azonban sokkal közelebb áll az emberi agy működéséhez. Kifejezetten idősorok vagy szekvenciák feldolgozására alkalmas és az előbb említett tulajdonságokat is képes kezelni. Kevésbé érzékeny a zajra [1], képes felügyelt vagy felügyelet nélküli tanulásra, az adatokat online dolgozza fel és képes lehet egy vagy kevés iterációból az adatsort megfelelően modellezni. A HTM pozitív tulajdonságai ellenére még nem terjedt el széleskörűen, ugyanis ennek a módszernek is vannak hiányosságai.

Kutatási célom volt a HTM implementációján javítani (a saját implementációm a hálózat neuronjai közötti ritka kapcsolatok optimalizálásán alapult [2]), annak teljesítményét és képességeit többféle szekvencián és idősoron megfigyelni. A vizsgált adatok között található szintetikus, valamint valódi szekvencia és idősor is. A megoldásomat egyrétegű LSTM hálózattal, valamint a Numenta csoport által implementált Nupic HTM hálózattal vetettem össze. A mérések a hálózatok gyorsaságát, a szükséges iterációk mennyiségét, valamint a predikciós pontosságukat tartalmazzák, és a teljesítmény fokozásának lehetőségeit tovább kutatom [3].

Referenciák

- [1] AHMAD SUBUTAI, JEFF HAWKINS: *Properties of sparse distributed representations and their application to hierarchical temporal memory*. arXiv preprint arXiv:1503.07469, 2015.
- [2] PILINSZKI-NAGY, CSONGOR: *Optimization of Hierarchical Temporal Memory through Efficient Sparse Representations*. BME VIK – TDK, Budapest, 2019.
- [3] PILINSZKI-NAGY, CSONGOR: *Optimization of Sparse Hierarchical Temporal Memory*. BME – MSc Diploma, Budapest, 2019 (szerkesztés alatt).

Ajánlati könyvek modellezése mély tanulóval

Hallgató: Reizinger Patrik

Mentor: Dr. Gyires-Tóth Bálint

Munkánk célja időbeli adatok elemzése volt mély neurális hálózatokkal annak érdekében, hogy a mesterséges intelligencia segítségével egyrészt pontosabb modellek kerüljenek megalkotásra, másrészt pedig, hogy ezen módszerek segítségével az adatrekordok által reprezentált folyamat látens terét minél jobban leképezzük és értelmezzük. A területet érintő munka egyik irányvonala a neurális hálózatok általánosítóképességének vizsgálata volt. A kutatás során kifejlesztésre került két, súlymátrix-alapú regularizáció technika [1][2], mely zajjal terhelt idősorok esetében szignifikáns javulást eredményezett. Továbbá a hagyományosnak számító hálózatok (LSTM, VAE) mellett a legújabb eredmények, mint például a Transformer [3] modellek is a vizsgálat tárgyát képezték. A szakirodalom által ismert architektúrák mellett azok továbbfejlesztésére is sor került, hogy a módszerekben rejlő flexibilitást kiaknázhassuk. Ennek során a Transformer architektúra átalakításra került, hogy az eredetileg nyelvfeldolgozáshoz kialakított hálózat idősor-elemzésre is alkalmazható legyen, továbbá egy speciális VAE hálózat is tárgyát képezte a kísérleteknek, ahol attention mechanizmus került alkalmazásra.

A vizsgálatok célpontjai elsődlegesen pénzügyi idősorok voltak, annak is egy speciális fajtája, az ajánlati könyv. Megmutattuk, hogy az alapmodellekhez képest a legújabb architektúrák jelentős javulást képesek elérni osztályozási feladatban, mindezt ráadásul technikai adatjellemzők használata nélkül, ráadásul kiegyenlítettlen osztályok esetében is.

Referenciák

- [1] REIZINGER, P. and GYIRES-TÓTH, B. (2019): *Stochastic Weight Matrix-based Regularization Methods for Deep Neural Networks*. 5th International Conference on Learning, Optimization and Data Science, Siena, Italy, September 10-13, 2019.
- [2] REIZINGER, P. (2019): *Súlymátrix-alapú, sztochasztikus regularizációs technikák vizsgálata mély neurális hálózatokban*. XXXIV. OTDK Informatika Tudományi Szekció 2019. Óbudai Egyetem.
- [3] Ashish VASWANI et al.: *Attention is all you need*. Advances in neural information processing systems. 2017. pp. 5998-6008.

Intelligens adatelemzés kiberfizikai (CPS) rendszerekben

Az adatmodellezésben, -értelmezésben és -feldolgozásban bonyolult elemzést igényel a nagy mennyiségű, heterogén, komplex adat, a fizikai környezettel és ezen belül a humán résztvevőkkel való interakció. Így a feldolgozási folyamat rendkívül összetett, megtervezése, áttekintése, automatizálása (a szenzorszinttől, a különböző absztrakciójú modelleken át a döntéstámogatásig, magyarázatgenerálásig) nélkülözhetetlené teszi az intelligens módszereket.

Az adott időszakon belül a sokrétű téma több területén folytak intenzív – hallgatókat is bevonó – kutatások. Az egyik terület a tárgyterületi ontológiával támogatott adatalapú szövegelemzés, ami a kiberfizikai rendszerekben megjelenő humán résztvevőkkel való kommunikáció miatt fontos. Másik terület az intelligens ágensmodelleken alapuló városi közlekedés vizsgálata. Nagy hangsúlyt nyert a mesterséges neurális hálók alkalmazási problémáinak kutatása, részben képfeldolgozási területen, másrészt a komplexitáscsökkentő módszerek kutatása. A kutatások eredményeképpen több szakdolgozat, diplomaterv és két konferenciacyikk született.

Intelligent Data Analysis in Cyber-Physical Systems (CPS)

Cyber-physical systems bring up new challenges, due to the huge amount of complex heterogeneous data and intensive interactions with the physical and human environment. This complexity makes the information processing also highly sophisticated as well. Therefore, the development, the control, and the automation of the processing call for intelligent (i.e., knowledge intensive) methods should be applied at several levels of abstraction, from the sensor level up to the decision support and automatic explanation generation.

Intensive research has been conducted in many areas. One research group has investigated ontology based text analysis, crucial for the mixed human-computer interaction in CPS. Another group has conducted research on intelligent agent based traffic in cities. Application of artificial neural networks was an important area as well. The theory of network complexity reduction; and application of neural networks to image processing problems have been investigated. As a result 2 thesis works, 2 conference papers were published in this period.



Dr. Dobrowiecki Tadeusz

Méréstechnika és Információs Rendszerek Tanszék
Department of Measurement and Information Systems
dobrowiecki@mit.bme.hu

Intelligens gépkocsik városi forgalomban – kommunikáción alapuló kooperáció vizsgálata

Hallgató: Alekszejenkó Levente

Mentor: Dr. Dobrowiecki Tadeusz

Az elmúlt évtizedekben tapasztalható közúti városi forgalomnövekedés új kihívások elé állította a mérnököket. Mivel a közutak szélesítése költséges és nem egy fenntartható fejlesztési módszer, ezért máshogy szükséges növelni a közutak kapacitását. Ilyen kapacitásnövelő módszer az egyes jelzőlámpával ellátott kereszteződések közötti jelzéskoordináció, azaz a forgalomirányító berendezések jelzéseinek valamilyen módon történő összehangolása annak érdekében, hogy a járműcsoportok minél rövidebb idő alatt elhagyhassák a koordinációban részt vevő csomópontokat. Manapság ennek az egyik legelterjedtebb módja a zöldhullámok kialakítása, illetve jelzőlámpák telepítése is. Mindkét megoldásnak vannak előnyei, de vannak hátrányai is.

Az önvezető járművek és az IoT-eszközök terjedésével lehetőség nyílik új megoldások létrehozására. Ilyen lehet egy jármű infrastruktúra (V2I) és az infrastruktúra elemei közötti kommunikáción alapuló új jelzéskoordinációs módszer. A megvalósítandó jelzéskoordinációs módszertől elvárjuk, hogy ne igényelje az artériák definícióját, valós idejű és a várakozási időt tekintve fair legyen, valamint jól skálázódjon a kereszteződések számával.

A konkrét kutatás célja egy ilyen tulajdonságú koordinációs módszer kialakítása és szimulációkkal segített elemzése. A szimulációs platform a már sikeresen alkalmazott és bővített SUMO mikroszkopikus forgalomszimulátor, illetve (szükség szerint) annak továbbfejlesztett változatai. A vizsgált forgalmi helyzet (a kutatás folytonossága és az eredmények összehasonlíthatósága érdekében) a budapesti BAH közlekedési csomópont topológiája.

Referenciák

- [1] ALEKSZEJENKÓ L. and T. DOBROWIECKI: *Intelligent Vehicles in Urban Traffic – Communication Based Cooperation*, the IEEE 17th World Symp. on Applied Machine Intell. and Inform. (SAMI 2019), Jan 24-26, 2019, Herľany, Slovakia.
- [2] ALEKSZEJENKÓ L.: *Intelligens gépkocsik városi forgalomban – Kommunikáción alapuló kooperáció vizsgálata*. BSc szakdolgozat, BME-VIK, 2018 (jeles).
- [3] ALEKSZEJENKÓ L. and T. DOBROWIECKI: *SUMO Based Platform for Cooperative Intelligent Automotive Agents*. In: Weber, M.; Bieker-Walz, L.; Hilbrich, R.; Behrisch, M. (szerk.) SUMO User Conference 2019, EasyChair Publishing (2019) pp. 107-123.

Neurális hálók komplexitásának automatizált redukciója

Hallgató: Formanek András

Mentor: Hadházi Dániel

Napjainkra számos mérnöki feladat megoldása során a neurális paradigma apparátusának alkalmazása alapértelmezetté vált. Ennek oka leginkább abban keresendő, hogy sok, aktívabban kutatott probléma (pl. döntéstámogatási, adatelemzési, képfeldolgozási feladatok) a konvencionális, algoritmusokon alapuló megoldások lehetőségeinek határait feszegeti. Az adaptív rendszerek elméletére alapozó neurális hálókkal eszközölt feladatmegoldás során elegendő megfelelő minőségű tanító mintakészlet biztosítása, melyből a hálózatok képesek tanulni. Ennek gyakorlati megvalósítása azonban számos tekintetben nehézkes. Egyrészt nem áll rendelkezésre megfelelő nagyságú és minőségű tanító mintakészlet (néhány benchmark feladatot leszámítva, mint az ImageNet, PascalVOC, stb. versenyek). Másrészt a gépi tanulás általánosító képességére vonatkozó elvárás, illetve a számítási erőforrásigény nem (vagy csak nehezen) teljesíthető.

A tématerület szakértői számára ismert a torzítás – variancia dilemma, az ERM konvergenciájának sebességére vonatkozó összefüggések, valamint Vapnik és Chervonenkis statisztikai tanulásmélete, melyek alapján egy hálónak a pontossága csak abban az esetben növelhető, ha növeljük a tanítóminták számát, vagy redukáljuk a szabad paramétereinek az effektív számát. Ez gyakran olyan regularizációkkal történik (pl. Tikhonov, Dropout, Batch Normalization stb.), mellyel a háló valódi paramétereinek a száma nem csökken, ezáltal azok éles működése sem gyorsul. A konkrét kutatás olyan eljárás kidolgozására fókuszál, mely a két fentebbi célt egyszerre próbálja elérni. Eddigi eredmények alapján a kidolgozott módszer hatékonyabb az általános nyesési eljárásoknál, melynek oka az eltérő megközelítésben gyökeredzik. Jelenleg is vizsgáljuk, hogy hogyan bővíthető a javasolt módszer alkalmazhatósága, az elért teljesítménynövekedés hogyan terjeszthető ki teljesen általános felépítésű hardware-ekre is. Mindemellett vizsgáljuk az eljárás egyéb paraméterezési lehetőségeit is, mellyel a jelenleginél direkter hangolható megoldást kaphatunk.

Referenciák

- [1] FORMANEK András: *Hatékony konvolúciós neurális hálózat tervezése osztályozási problémákra*. TDK dolgozat BME-VIK, 2018. november, 1. díj.
- [2] FORMANEK András: *Designing effective convolutional neural network for processing image information*. BSc szakdolgozat, BME-VIK 2018.
- [3] FORMANEK András: *Hatékony konvolúciós neurális hálózat tervezése osztályozási problémákra*. OTDK dolgozat Informatika Tudományi Szekció, 2019. április, különdíj.

Képszegmentálás városi környezetre jellemző objektumok detektálására önvezető járművekhez

Hallgató: Révy Gábor

Mentor: Dr. Hullám Gábor

A városi környezetre jellemző objektumok felismerése egy lényeges komponensévé válhat a jelenleg is dinamikusan fejlődő vezető támogató rendszereknek, illetve későbbiekben az önvezető járművek és autonóm robotok megfelelő rendszereinek. Városi környezetben szereplő objektumok, úgymint gyalogos, járda, út, bicikli, autó vagy közlekedési tábla felismerése a járműre szerelt kamera által biztosított felvétel feldolgozása révén lehetséges. Ennek egy módja a képszegmentálás, amely során egy képet úgy osztunk részekre, hogy egy rész egy adott objektumtípushoz tartozzon. Ezt követően minden egyes pixelhez rendelünk egy címkét, amely megadja, hogy az adott pixel milyen típusú objektumhoz tartozik. Az objektumok meghatározása pedig azért lényeges, mert ennek függvényében célszerű befolyásolni az adott vezetési módot, például figyelmeztetni a vezetőt egy oldalról érkező kerékpárról, vagy adott esetben beavatkozni, ha egy gyalogos túl közel kerül a várható útvonalhoz. Ezen feladat automatizált megoldása jelentős előnyt jelentene, főleg, ha az algoritmus megfelelően pontos és kellően gyors.

A kutatás célja egy képszegmentálásra alkalmas mély neurális háló alapú rendszer kialakítása. Ennek részeként vizsgáltunk meg különböző architektúrájú konvolúciós neurális hálókat [1], különös tekintettel a dilátált konvolúciót tartalmazó megoldásokra [2]. Ezek figyelembevételével terveztünk meg [3] több lehetséges mély neurális hálót, melyek tanításához olyan valós városi környezetből származó felvételeket tartalmazó adatbázist használunk fel, mint például a tématerületen ismert Cityscapes adatbázist. A tanítás előkészítése során figyelembe kell venni a különböző adathalmazokban elérhető objektumok típusait, az elérhető tanítóminták számosságát, továbbá a rendelkezésre álló képek minőségét. Tehát több feldolgozási lépést követően áll elő a megfelelően címkézett tanítóminták halmaza. A tanítás eredményességét többféle mutatóval értékeljük ki, az egyes objektumtípusoknál fellépő esetleges hibák jellemzőit részletesen megvizsgáljuk.

Referenciák

- [1] Y. GUO, Y. LIU, T. GEORGIU and M. S. LEW: *A review of semantic segmentation using deep neural networks*. International journal of multimedia information retrieval, 7.2, pp. 87–93, 2018.
- [2] L.-C. CHEN, Y. ZHU, G. PAPANDREOU, F. SCHROFF and H. ADAM: *Encoder-decoder with atrous separable convolution for semantic image segmentation*, in *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 801–818.
- [3] RÉVY GÁBOR: *Képszegmentálás városi környezetre jellemző objektumok detektálására*. TDK dolgozat, BME, VIK, 2019.

Nagy adathalmazok (Big Data) komplex struktúráinak előrejelzése az adattudomány módszereivel

A mindennapokban keletkező nagy mennyiségű strukturálatlan adat feldolgozása komplexitásából és méretéből kifolyólag hagyományos eszközökkel már nehezen valósítható meg. Ezért egyre nagyobb hangsúly helyeződik az adattudomány és a mesterséges intelligencia korszerű módszereire, melyek megkönnyítik az adatokban rejlő összefüggések kinyerését. E területen 3 téma kutatásával foglalkoztunk: a természetes nyelvfeldolgozáshoz tartozó összefoglaló generálással, komplex struktúrájú adathalmazok osztályozásával, klaszterezésével, valamint implicit visszacsatoláson alapuló kollaboratív szűrős ajánlórendszerrel. A kutatott módszerek között a mély neurális háló volt az, mely szinte mindegyik témánál előfordult, hiszen napjainkban ennek az adaptálhatósága és hatékonysága a legnagyobb. Mély neurális hálóval oldottuk meg a szövegdokumentumok összefoglalóinak automatikus előállítását és az ajánlórendszereknél is használtuk, ahol a mátrix faktorizációt vetettük össze vele. Több TDK dolgozat és 2 nemzetközi konferenciacikk született a kutatásunk eredményeként.

Predict the complex structures of large data sets with the methods of data science

Due to its complexity and size, the processing of large amounts of unstructured data in everyday life is difficult to achieve by conventional tools. Therefore, there is an increasing emphasis on advanced methods in data science and artificial intelligence, which facilitate the extraction of hidden relationships in data. In this area, we dealt with three topics: (i) summary generation belonging to natural language processing, (ii) classification, clustering of complex structured data sets, and (iii) recommendation system with collaborative filtering based on implicit feedback. Among the researched methods the deep neural network occurred in almost every topic, since its efficiency and adaptability are the highest nowadays. Deep neural network was used to solve automatic summary generation of text document and was also used in recommendation systems where matrix factorization was compared with it. As a result of the research, more TDK papers and 2 international conference articles were produced.



Témavezető: Dr. Szűcs Gábor

Távközlési és Médiainformatikai Tanszék

Department of Telecommunications and Media Informatics

szucs@tmit.bme.hu

Szövegdokumentumok összefoglalóinak automatikus előállítás

Hallgató: Huszti Dorottya

Mentor: Dr. Szűcs Gábor

A természetes nyelvfeldolgozáshoz tartozó kutatásunk célja automatikus összefoglaló generálást végző rendszer megtervezése és megvalósítása volt, ugyanis napjaink egyik legnagyobb kihívása a természetes nyelvű dokumentumok tartalmának rövid, de pontos összefoglalása, melyet kivonatolásnak is hívnak. Ez egy olyan összetett feladat, mely a szöveg megértését és a releváns információk kinyerését egyidejűleg igényli. A feladat megoldásának két megközelítése létezik, így két nagy típust tudunk megkülönböztetni: egy absztrakt (abstractive) és egy kinyerés (extractive) alapú összefoglaló generáló rendszert. A kutatás elején kielemeztük a különböző szakirodalmi módszereket, valamint a feladathoz felhasználható technológiákat. Ezt követően részletesen megterveztuk a gépi tanuláson (mély neurális hálón) alapuló összefoglaló generáló rendszer felépítését. A felépítésnél olyan mély neurális hálózat architektúrát javasoltunk az absztrakciós összefoglaláshoz, mely egy kódoló és dekódoló részből állva el tudja majd készíteni a rövid összefoglalókat a hosszú szöveges dokumentumokból. A hagyományos architektúrák egyirányú kódolóval rendelkeznek, a javasolt megoldásunknak viszont olyan kétirányú kódolója van, amely 2 darab LSTM (Long Short-Term Memory) hálózatot tartalmaz. Az egyik LSTM a mondatok elejétől, a másik pedig a végétől kezdi a feldolgozást, a végső állapotot pedig az LSTM-ek belső állapotainak kombinálásával hozza létre. Ezen kívül kétféle dekódert használtunk, egyet a tanításhoz, és egyet a predikcióhoz. Ez utóbbi a beamsearch ötletén alapszik, amely egynél több jelöltet is megtart, amikor szekvenciaelemeket generál. Ennek a dekódernek a végső becslése a legnagyobb valószínűségű sorrend lesz. A kísérleteket a Newsroom adathalmazon mutattuk be, mely olyan párokat tartalmazott, ahol a pár egyik tagja a hosszú szöveges dokumentum, másik pedig a hozzá tartozó rövid összefoglaló volt. A kutatás eredményeiből egy friss TDK dolgozat és egy nemzetközi IEEE konferencia cikk készült.

Referenciák

- [1] Huszti Dorottya: *Szövegdokumentumok összefoglalóinak automatikus előállítás*. TDK dolgozat, BME, VIK, 2019.
- [2] Szűcs, G., Huszti, D.: *Seq2seq Deep Learning Method for Summary Generation by LSTM with Two-way Encoder and Beam Search Decoder*. IEEE 17th International Symposium on Intelligent Systems and Informatics Proceedings, SISY 2019, Subotica, Serbia, September 12-14, 2019, pp. 221-225.

Aktív tanulás és spektrális klaszterezés nagy adathalmazoknál

Hallgató: Knoll Zsolt

Mentor: Papp Dávid

Témánk fókuszában a mesterséges intelligencia aktív tanulási ága állt, melynek lényege, hogy csak kevés címkézett példány áll a tanuló algoritmus rendelkezésére, de egy intelligens lekérdezési stratégia segítségével választhat a címkézettlen példányok közül. Kutatásunk során két aktív tanulási módszert javasoltunk, melyeknél mindkettő egy-egy különbségszámítási ötletre épül. Az egyik új módszer a különbségi mintavételi lekérdezési stratégia: ez a stratégia kiszámítja az új különbséglistát, melynek elemei befolyásolni tudják az osztálycímkével nem rendelkező példányok bizonytalansági mértékét. A különbségek listáját figyelembe véve szeretnénk lekérdezni a legkisebb különbségi értékkel rendelkező példányt; másrészt, csak a bizonytalansági értékek alapján (itt a maximális entrópia kritériumokat használtuk) a maximális értékre törekszünk. Ezen mutatók arányának kiszámítása után a különbségi mintavételi stratégia célja ennek az arálynak a minimalizálása (minimális számláló, maximális nevező). A másik új módszer a várható különbségváltozási lekérdezési stratégia, mely úgy nevezett globális különbségi mutatóval rendelkezik az egyes lépések számára. Ez a stratégia egyesíti a várható modellváltozás és a bizonytalanság mintavétel stratégiáit, figyelembe véve a bizonytalansági értékek különbségét. Ennek a kombinált stratégiának a célja egy olyan példány osztálycímkéjének a lekérdezése, amely valószínűleg a következő lépés globális különbségének legnagyobb változását eredményezi. A kísérleti eredmények azt mutatták, hogy mindkét különbségen alapuló mintavételi lekérdezési stratégia magasabb jósági értékeket ért el a szakirodalomban található frissen publikált módszerekhez képest.

Ezen kívül klaszterezéssel foglalkoztunk, melynek létezik 2 típusa: kemény és puha klaszterezés. A kemény klaszterezés csoportjába azok a módszerek tartoznak, melyeknél egy pont vagy beletartozik egy klaszterbe, vagy nem. Ezzel ellentétben a puha klaszterezésnél minden klaszternek része valamilyen mértékben az adott pont. A témában egy TDK dolgozat készült, melynek fő célja az volt, hogy a klaszterező algoritmus hierarchikus struktúrájú (azaz ahol a ponthalmazok még magasabb szinten lévő ponthalmazokba vannak szervezve) adatokon is alkalmazható legyen. Mind az osztályozás (aktív tanulással), mind a klaszterezés gyakorlati szempontból is hasznos, kipróbáltuk őket különböző nagy adathalmazokon.

Referenciák

- [1] KNOLL Zsolt: *Hierarchikus struktúrájú adathalmazok spektrális klaszterezése*. TDK dolgozat, BME, VIK, 2019.
- [2] PAPP, D., SZÜCS, G., KNOLL, Zs.: *Difference based query strategies in active learning*. IEEE 17th International Symposium on Intelligent Systems and Informatics Proceedings, SISY 2019, Subotica, Serbia, September 12-14, 2019, pp. 35-39.

Implicit visszacsatoláson alapuló kollaboratív szűrős ajánlórendszer mély neurális hálóval

Hallgatók: Kiss Richárd, Pogány Domonkos, Sándor Dániel

Mentor: Dr. Szűcs Gábor

A mai világban a jelentős adatnövekedésnek és a szolgáltatók közötti versenynek köszönhetően megnőtt az igény az egyre jobb teljesítményű ajánlórendszerek kifejlesztésére; ezeknek a rendszereknek a célja, hogy minél relevánsabb ajánlatokat adjon a felhasználóknak. Az ajánlórendszerek szakirodalmában alapvetően kétféle csoportosítás létezik. Az egyiknél az adatok származása alapján történik csoportosítás, itt megkülönböztetünk implicit (amikor a felhasználó és a termék közti interakcióból automatikusan származnak az adatok) és explicit (amikor egy felhasználó a termék fogyasztása után értékeli a terméket) értékeléseket, visszajelzéseket. Másik csoportosításnál pedig a termékekről rendelkezésre álló információk alapján választhatjuk szét a két ajánlórendszer típust: tartalom alapú szűrés, ahol a termékekhez plusz információk is vannak, illetve a kollaboratív szűrés, mely matematikai konstrukciók segítségével látens tulajdonságokat rendel a termékekhez, illetve felhasználókhoz. Kutatásunkban az utóbbival, illetve az implicit visszajelzés típusú ajánlásokkal foglalkoztunk részletesen. Kezdetben az úgynevezett baseline módszert használták az iparban, azaz minden felhasználónak egységesen a legjobb termékeket ajánlották. Ezután terjedtek el a hasonlóság alapú módszerek, amelyek a felhasználóknak egyedi ajánlásokat tettek, a hozzájuk hasonló felhasználók, illetve az általuk használt termékek alapján. Az első komolyabb előrehaladást a témában a mátrixfaktorizációs modellek jelentették, amelyek a felhasználókat és a termékeket is ugyanabban a látens térben reprezentálják, majd ezek cosinus távolsága alapján rangsorolják a releváns termékeket. A legfrissebb kutatások arra irányulnak, hogy neurális hálóval végezzék a látens vektorok hozzárendelését. Kutatásunk során megvizsgáltuk a régebbi ajánlórendszereket és a legfrissebb módszereket, elemeztük az eltéréseket. Különböző pontosság metrikák alapján összehasonlítottuk és kombináltuk őket egy saját ajánlórendszer létrehozása érdekében, majd kiértékeljük a kapott eredményeket. Ebben az implicit visszacsatoláson alapuló kollaboratív szűrős ajánlórendszer témában végül egy TDK dolgozat készült, melyben mély neurális hálót használtunk a jobb ajánlások elérése érdekében.

Referencia

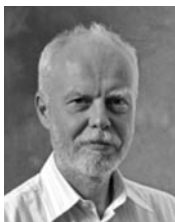
- [1] KISS Richárd, POGÁNY Domonkos, SÁNDOR Dániel: *Implicit visszacsatoláson alapuló kollaboratív szűrős ajánlórendszer mély neurális hálóval*. TDK dolgozat, BME, VIK, 2019.

Döntéstámogató módszerek dinamikus orvosi képalkotó berendezésekhez

Az elmúlt néhány évtizedben a 3D információt előállító képalkotó berendezések alapvető diagnosztikai eszközzé váltak az orvosi gyakorlatban az egészségügy számos területén. Napjaink képalkotó berendezései azonban már egy-egy fizikai változót nemcsak térben, de időben is képesek rekonstruálni, azaz 4D adatot állítanak elő. A hosszú mérési idő miatt, a páciensek elmozdulása jelentős zajokat visz be a mérésbe, így ennek kompenzálása kiemelten fontos. A keletkezett adatmennyiség egyetlen mérés esetén is óriási. Az adatok feldolgozása, a rejtett összefüggések kinyerése Big Data eljárások alkalmazását igényli. Célunk a klasszikus képfeldolgozó módszerek mellett mélytanulás alkalmazása, mely hatékonyan képes figyelembe venni a 4D adatok sajátosságait. A hatalmas adatmennyiség elfogadható időn belüli feldolgozása megköveteli a párhuzamosított, GPU alapú implementációt. Az új módszerek elsősorban orvosbiológiai és gyógyszerészeti kutatásokban alkalmazhatóak.

Decision support methods for dynamic medical imaging systems

In the last few decades, 3D medical imaging devices became primary source of diagnostic information in many areas of the clinical practice. Current imaging devices are capable to measure physical properties not just in spatial but in temporal dimensions as well. This makes it possible to collect 4D information. Unfortunately, due to the long exposure time, patient movements introduce significant noise in the measurement that has to be compensated. The amount of the obtained data is huge for a single measurement. To process the information and to reveal hidden dependencies, Big Data methods shall be used. Our goal is to apply not just the framework of the classic image processing but to extend deep learning methods to take the properties of the 4D medical imaging devices into consideration to be able to provide efficient solution while utilizing the parallel processing capability of the GPU. The new results would be definitely welcome in biomedical and pharmaceutical research.



Témavezető: Dr. Szirmay-Kalos László

Irányítástechnika és Informatika Tanszék

Department of Control Engineering and Information Technology

email: szirmay@iit.bme.hu

Neurális hálózattal támogatott PET rekonstrukció szinogrammm adatokból

Hallgató: Rácz Gergely Ferenc

Mentorok: Tóth Márton, Kacsó Ágota Enikő

A tomográfias rekonstrukció az orvosi képalkotás alapvető módszere, amellyel a mérőberendezésekből származó adatokból az orvosi vizsgálatokra alkalmas képi adatok állíthatók elő. Egy gyakran alkalmazott képalkotó eljárás a PET. A rekonstrukció az esetek többségében időigényes, bonyolult feladat, melyet iteratív algoritmusokkal oldanak meg. A legfrissebb kutatások szerint a gépi tanulás eszközeivel a rekonstrukció bizonyos részei a hagyományos módszereknél hatékonyabban hajthatók végre. Jelen projekt keretében azt vizsgáljuk, hogy mély neurális hálózatok segítségével a PET rekonstrukció mely lépéseit lehet javítani. Kutatócsoportunkkal először egy olyan zajszűrő neurális hálózatot fejlesztettünk, amely az iteratív algoritmus által számított zajos képek jel/zaj arányát növeli az anatómiai régiók határainak figyelembevételével. A kimeneti felvételek minőségét nagyban befolyásolja a PET rekonstrukcióban használt kompartment modell kinetikus paramétereinek megfelelő kezdeti becslése, ezért a zajszűrő hálózat mellett egy paraméterbecslő eljárás fejlesztésébe kezdtünk. A modellparaméterek gyors és pontos becslésére többretegű perceptront alkalmaztunk, mellyel sikerült a hagyományos iteratív globális optimalizálókkal hasonló hatásfokú, ám jóval gyorsabb becslést adni.

A kezdeti paraméterbecslő, valamint a zajszűrő hálózattal célunk a hagyományos iteratív rekonstrukciós algoritmus kiegészítése volt a számítást gyorsító, illetve korrigáló lépésekkel. Csoportunkkal jelenleg egy olyan rendszeren dolgozunk, amely a teljes rekonstrukciót képes a gépi tanulás eszközeivel megvalósítani. Ennek a központi eleme egy mély neurális háló, melynek bemenetére közvetlenül a szinogrammot csatoljuk, és a kimenetén megkapjuk a rekonstrukció eredményét.

Referenciák

- [1] TÓTH, B., TÓTH, M., KACSÓ, Á., RÁCZ, G. F. & SZIRMAY-KALOS L. (2018): *Controlling TV Regularization with Deep Learning*. 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), Sydney, Australia, pp. 1-5.
- [2] RÁCZ, G. F., KACSÓ, Á., TÓTH, M., & TÓTH, B. (2019): *Fast and accurate initial parameter fitting in Positron Emission Tomography using Neural Networks*. Proceedings of the Workshop on the Advances of Information Technology: WAIT 2019.

PET rekonstrukció javítása szűréssel és mozgáskompenzációval

Hallgató: Varnyú Dóra

Mentor: Kacsó Ágota

A PET képképzés legnagyobb kihívását a rekonstruált aktivitásértékek zajossága adja, amely kiküszöbölését regularizációval lehetséges megoldani. A regularizációt a sziták módszerével is meg lehet valósítani. A módszer lényege, hogy a rekonstrukció minden iterációjában az aktivitások aktuális becslését egy előre meghatározott szűrő kernellel konvolváljuk, ezáltal elsimítjuk benne a kiugró értékeket.

A projekt keretében számos szűrési módszert megvizsgáltunk, úgy mint a Gauss-szűrő, bilaterális szűrő, spike szűrő, medián szűrő és az irányított (guided) szűrő. Ezek közül az irányított szűrés vezetett a legjobb eredményre. Az irányított szűrő élmegőrző simítást végez el a bemeneten egy második, az ún. irányító kép segítségével. Az irányító lehet maga a bemenet vagy egy másik kép is, jelen esetben a PET felvétel mellett készített CT kép. Az egyes kimeneti értékek számolásakor a szomszédos értékek mellett figyelembe veszi az irányító megfelelő térbeli szomszédságának statisztikai értékeit is. Ha az irányító eltér a bemenettől, akkor az irányító struktúrája befolyással lesz a kimenő, szűrt kép struktúrájára. Ezt a hatást struktúra átvitelnek nevezik. A kutatási másik feladata a mozgáskompenzáció vizsgálata volt, amely figyelembe veszi, hogy a páciens nem mozdulatlan a vizsgálat során. A páciens mozgása különösen neurológiai vizsgálatoknál elkerülhetetlen. A maximum-likelihood elv felhasználásával általánosítottuk a PET rekonstrukció statikus esetre kidolgozott ML-EM iterációs sémáját, és megállapítottuk, hogy a mozgáskompenzációhoz a fizikai szimuláció előtt végrehajtott mozgás alapú elmosása, valamint a korrekció során két fordított elmosás szükséges.

Referenciák

- [1] VARNYÚ Dóra, SZIRMAY-KALOS László: *3D Motion Blur for PET Using GPGPU*. Central European Seminar on Computer Graphics (CESCG), Smolenice, 2019.
- [2] VARNYÚ Dóra, SZIRMAY-KALOS László: *Thick Line Integration with Filtered Sampling*. Képfeldolgozók és Alakfelismerők Konferenciája, Debrecen, 2019.
- [3] SZIRMAY-KALOS László, VARNYÚ Dóra: *Model-based Motion Compensation for Dynamic PET*. Képfeldolgozók és Alakfelismerők Konferenciája, Debrecen, 2019.

