OTKA FK-131793 project
Scientific report
2019.12.01-2022.11.30

Title:

Analysing IoT device and application behaviour modelling
to foster efficient management of Cloud and Fog environments

Principal Investigator:

Attila Kertesz, Ph.D.
University of Szeged

Submission Date: 2022.12.14.

Contents

1 Summary

In our project, we conducted research on the topic of simulation and behaviour analysis of IoT devices and applications. We progressed according to the work plan throughout the implementation of the project. At the beginning, we performed a detailed literature review, and then developed a general IoT behaviour model.

We extended the DISSECT-CF simulation environment by Fog and IoT infrastructure, mobility, localisation and IoT workflow modelling, trace loading and graphical interface functions. With our extended framework, we modelled and analysed the behaviour of several use cases, using real sensor data. We developed and analysed data management algorithms based on a fuzzy method, and developed new methods to increase the execution efficiency of workflow and MapReduce applications, and to enable data protection compliance and privacy protection.

We have achieved the most significant new results going beyond the state-of-the-art with the following contributions:

- *development of a behaviour model of IoT sensor and actuator devices and applications;*
- *simulation solutions for modelling the management of heterogeneous cloud, fog, dew and blockchain infrastructures;*
- *development of efficient and secure IoT data management methods.*

We carried out our evaluations with the involvement of several BSc and MSc students. A total of 27 publications were published within the framework of the project, which are publicly available and can be read on our ResearchGate page:

`https://www.researchgate.net/project/Analysing-IoT-device-and-application-behaviour-modelling-to-foster-efficient-management-of-Cloud-and-Fog-environments`

2 Detailed research results

According to our research plan, our aim in this project was to investigate the efficient and secure (privacy-aware) management of IoT-Fog-Cloud systems using behaviour models of IoT devices, sensors and applications, by performing fundamental research in the area of Fog and Cloud Computing and the Internet of Things (IoT). The main goals of our project were the followings:

- to gather, classify and make available real-world sensor datasets for investigating IoT applications and systems in simulated environments;
- to model the behaviour of IoT devices and sensors in various application fields,
- to demonstrate the wide applicability of these device models by realising, analysing and using them in different simulation environments for IoT-Fog-Cloud systems,
- to develop algorithms for analyzing the efficiency and privacy of sensor data processing, aggregation and storage with smart, fuzzy-based selection of fog and cloud gateway services.

In the following subsections we detail, how we have managed to reach these goals with our published papers.

2.1 Scientific results in the 1st year

We started our research for IoT device and application behaviour analysis by conducting a detailed literature search in the field of IoT simulation solutions supported by cloud and fog systems, and on the development challenges and applications of smart cities. We implemented a fog model to the DISSECT-CF simulator, and created the DISSECT-CF-Fog simulation environment, and analysed different scenarios of a smart region application (i.e. weather monitoring and forecasting). We also carried out further improvements and extensions of the simulator with the involvement of BSc students, in terms of localization and graphical interface functions. In addition to these, we also launched other, related directions, such as the examination of the execution efficiency of workflow and MapReduce applications, as well as examination of fog and blockchain simulation methods.

We held our first annual meeting on September 11, 2020, where all project participants presented their research progress and results. In the project proposal we planned 2 international conference presentations as scientific results for the first year. We managed to significantly overperform this plan by publishing 6 journal articles and 7 international conference papers.

Research tasks in the proposal for the first year, their corresponding publications and the summary of the main results of the papers:

- Task 1.1 Gathering real world IoT traces and analyzing them: [J1], [J2], [J3], [C4], [C5]
- Task 1.2 Examining real IoT application behaviour in Fog and Cloud environments: [C1], [C2], [C3], [C7], [J6]
- Task 1.3 Designing a generic IoT device and application behaviour model: [J4], [J5], [C5]
- Task 1.4 Implementing the behaviour model to the DISSECT-CF simulator: [C6]
- Task 1.5 Implementing the behaviour model to the MobIoTSim simulator: [D1]

[J1] Markus, A., & Kertesz, A. (2020). A survey and taxonomy of simulation environments modelling fog computing. SIMULATION MODELLING PRACTICE AND THEORY, 101. `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31038965`

IoT-Fog-Cloud systems are very complex, and the use of simulations in their design, development and operational processes is inevitable. There are many simulator solutions available to model and analyse these systems depending our research needs, but in many cases it is hard to grasp their differences, and implementing certain scenarios in different tools is time consuming. In this work we proposed a survey and taxonomy of the available simulators modelling clouds, IoT and specifically fogs. The main contributions of this study are our novel viewpoints for classification including software quality, which is performed by analysing the source code of the considered simulators. We also proposed comparison tables for three groups of simulators that reveal their differences and the way they model the elements of these systems. Finally, we discussed the relevant findings of our classifications, and highlighted open issues that need further research.

[J2] Baniata, H., & Kertesz, A. (2020a). A Survey on Blockchain-Fog Integration Approaches. IEEE ACCESS, 8, 102657–102668. `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31361709`

Fog computing (FC) is the extension of Cloud Computing (CC), from the core of the internet architecture to the edge of the network, with the aim to perform processes closer to end-users. This extension is proven to enhance security, and to reduce latency and energy consumption. Blockchain (BC), on the other hand, is the base technology behind crypto-currencies, yet is implemented in wide range of different applications. The security and reliability, along with the distributed trust management criteria proposed in BC, represent a step towards reaching a distributed and trusted management systems. In this survey paper we presented a detailed literature review and classification in FC-BC integration possibilities. We discussed and categorized the related papers according to the year of publication, domain,

used algorithms, BC roles, and the placement of the BC in the FC architecture. Our research presented detailed observations, analysis, and open challenges for the BC-FC integration.

[J3] Kirimtat, A., Krejcar, O., Kertesz, A., & Tasgetiren, M. F. (2020). Future Trends and Current State of Smart City Concepts: A Survey. IEEE ACCESS, vol. 8, 86448–86467. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31348900

Intelligent systems are wanting for cities to cope with limited spaces and resources across the world. Smart cities emerged mainly as a result of highly innovative ICT industries and markets, and additionally, they have started to use novel solutions taking advantage of the Internet of Things (IoT), big data and cloud computing technologies. The concept of a floating smart city emerged as a novel solution due to rising sea levels and land scarcity in order to provide alternative living spaces for humanity. In this paper we reviewed the current state of smart city concepts across the world by understanding the key future trends, including floating cities, by motivating researchers and scientists through new IoT technologies and applications. Our detailed literature survey and review applied a complex literature matrix including terms, like smart people, smart economy, smart governance, smart mobility, smart environment, and smart living. With the proposed approach, recent advances and practical future opportunities for smart cities can be revealed.

[J4] Baniata, H., Sharieh, A., Mahmood, S., & Kertesz, A. (2020). GRAFT: A Model for Evaluating Actuator Systems in terms of Force Production. SENSORS, 20(7). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31268910

In the scope of evaluation methodologies for Internet of Things (IoT) systems, some approaches concern security, while others latency. However, some methodologies evaluate systems that contain active entities, so-called actuators. In this paper we proposed a novel methodology for evaluating such systems with actuator components using Graph Representation of the Angle of the Force and Time called GRAFT. It facilitates easy computation of the net force produced by physical or mechanical acts occurring on a daily basis on Earth. We used laws and definitions of physics describing the relations between Speed, Distance, and Time, and applied them in a heliocentric system, and modeled the considered systems with a graph. We compared our default GRAFT algorithm to a special implementation using the Clock-Angle-Problem (CAP) for sessions. We used an IoT-focused use case for validating our approach, and we presented a detailed explanation of the proposed GRAFT algorithm. The experimental results showed the ability of GRAFT to provide highly accurate results, which also exemplifies that our GRAFT approach is programmable, hence deployable in real life scenarios.

[J5] Mishra, B., & Kertesz, A. (2020). The Use of MQTT in M2M and IoT Systems: A

Survey. IEEE ACCESS, 8, 201071–201086. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31674040

Nowadays billions of smart devices or things are present in Internet of Things (IoT) environments, such as homes, hospitals, factories, and vehicles, all around the world. As a result, the number of interconnected devices is continuously and rapidly growing. In this paper we analyzed the growth of M2M protocol research (including MQTT, AMQP, and CoAP) over the past 20 years, and showed how the growth in MQTT research stands out from the rest. We also gather relevant application areas of MQTT, as the most widespread M2M/IoT protocol, by performing a detailed literature search in major digital research archives. Our quantitative evaluation presented some of the important MQTT-related studies, which we compared to discuss the main features, advantages, and limitations of the MQTT protocol.

[J6] Baniata, H., Anaqreh, A., & Kertesz, A. (2021). PF-BTS: A Privacy-Aware Fog-enhanced Blockchain-assisted task scheduling. INFORMATION PROCESSING & MANAGEMENT, 58(1). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31623462

In complex usage scenarios, such as a smart city applications, where there is a large number of virtual resources (VR) to be managed, task scheduling is exposed as an NP-Hard problem. As there are many automated scheduling solutions proposed so far, new possibilities arise with the advent of Fog Computing (FC) and Blockchain (BC) technologies. In this paper we proposed an Ant Colony Optimization (ACO) algorithm in a Fog-enabled Blockchain-assisted scheduling model, namely PF-BTS. The protocol and algorithms of PF-BTS exploit BC miners for generating efficient assignment of tasks to be performed in the cloud's VRs using ACO, and award miner nodes for their contribution in generating the best schedule. PF-BTS also allows the fog to process, manage, and perform the tasks to enhance latency measures. While this processing and managing is taking place, the fog is enforced to respect the privacy of system components, and assure that data, location, identity, and usage information are not exposed. We evaluated PF-BTS performance with a recently proposed Blockchain-based task scheduling protocol in a simulated environment. Our evaluation and experiments showed high privacy awareness of PF-BTS, along with noticeable enhancement in execution time and network load reduction.

[C1] Markus, A., Gacsi, P., & Kertesz, A. (2020). Develop or Dissipate Fogs? Evaluating an IoT Application in Fog and Cloud Simulations. In Proceedings of the 10th International Conference on Cloud Computing and Services Science (pp. 193–203). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31338293

Clouds and fogs have promising properties to serve IoT needs, which require enormous data to be stored, processed and analysed generated by their sensors and devices. Since such IoT-Fog-Cloud systems can be very complex, it is inevitable to use simulators to investigate them. Cloud simulation is highly studied by now,

and solutions offering fog modelling capabilities have also started to appear. In this paper we briefly compared fog modelling approaches of simulators, and presented detailed evaluations in two of them to show the effects of utilizing fog resources over cloud ones to execute IoT applications. We also shared our experiences in working with these simulators to help researchers and practitioners, who aim to perform future research in this field.

[C2] Baniata, H., & Kertesz, A. (2020). PF-BVM: A Privacy-aware fog-enhanced blockchain validation mechanism. In Proceedings of the 10th International Conference on Cloud Computing and Services Science (pp. 430–439). `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31391880`

Fog computing is one of the recently emerged paradigms that needs to be improved to serve Internet of Things (IoT) environments of the future. In this paper we proposed PF-BVM, a Privacy-aware Fog-enhanced Blockchain Validation Mechanism, that aims to support the integration of IoT, Fog Computing, and the blockchain technology. In this model the more trusted a fog node is, the higher the authority granted to validate a block on behalf of the blockchain nodes. To guarantee the privacy-awareness in PF-BVM, we use a blockchain-based PKI architecture that is able to provide higher anonymity levels, while maintaining the decentralization property of a blockchain system. We also proposed a concept for measuring reliability levels of blockchain systems. We validated our proposed approach in terms of execution time and energy consumption in a simulated environment. We compared PF-BVM to the currently used validation mechanism in the Proof-of-Work (PoW) consensus algorithm, and found that PF-BVM can effectively reduce the total validation time and total energy consumption of an IoT-Fog-Blockchain system.

[C3] Baniata, H., Almobaideen, W., & Kertesz, A. (2020). A Privacy Preserving Model for Fog-enabled MCC systems using 5G Connection. In 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC) (pp. 223–230). `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31623473`

Privacy issues in Cloud Computing are one of the major concerns for many individuals and companies around the world, and still a challenge for research and industry. The appearance of the 5G technology and the latest advances in Mobile Cloud Computing gave way to fog-enabled systems that represent the future of current cloud systems. Accordingly, the Fog concept had been proposed in 2013 to enhance the IoT-Cloud operations in terms of latency and reliability. In this paper we analyzed and categorized surveys addressing privacy issues of these complex systems, and proposed PFMCC, a model for preserving Privacy in Fog-enabled Mobile Cloud Computing systems using 5G connection. The PFMCC model consists of three components interacting with each other to preserve data privacy, usage privacy, location privacy and high mobility. We also presented three algorithms to perform privacy-aware computation offloading to the fog.

[C4] Al-Haboobi, A., & Kecskeméti, G. (2020). Reducing Execution Time of An Existing Lambda based Scientific Workflow System. In The 12th Conference of PhD Students in Computer Science - Volume of short papers (pp. 3–6). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31383116

Recent developments in the field of cloud computing have led to emerging an innovative service called Function as a Service (FaaS). Cloud functions have received increased attention across a number of disciplines in recent years. Researchers have executed scientific applications such as workflows using function platforms. Functions could reduce the execution time and cost for scientific workflows due to their features: auto-scaling and fine-grained pricing schema. In this paper we developed an improved system based on the DEWE v3 system that can process large-scale workflows with three different execution modes: (virtual machines, cloud functions, and a hybrid mode). We can improve DEWE v3 by modifying its scheduling algorithm in the cloud function environment. The improved algorithm can schedule a group of jobs with precedence requirements to run in a single function invocation for speeding up the execution time. We evaluated the improved system with large-scale Montage workflows by comparing its result with the original DEWE v3. The experimental results showed that the improved system can minimize the execution time and cost in contrast to DEWE v3 in most cases.

[C5] Ebenezer, K. G., & Kecskeméti, G. (2020). Improving MapReduce Speculative Executions with Global Snapshots. In The 12th Conference of PhD Students in Computer Science - Volume of short papers (pp. 62–65). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31366543

Hadoop is a MapReduce implementation for distributed storage and computation, which has issues with managing poor performing jobs. A challenge called speculative execution, is mostly handled by running backup tasks. In this paper we proposed an application of consistent global snapshots and stable property to resolve this challenge. This involves the capturing of snapshots of all data I/O into mappers and reducers before and after data executions. The snapshots are then compared to determine the poor performing tasks. These tasks are quickly divided and redistributed amongst the inactive mappers and reducers based on an algorithm on data complexity.

[C6] Sallo, D. H., & Kecskeméti, G. (2020). Parallel Simulation for The Event System of DISSECT-CF. In The 12th Conference of PhD Students in Computer Science - Volume of short papers (pp. 58–61). https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31383115

Discrete Event Simulation (DES) frameworks gained significant popularity to support and evaluate cloud computing environments, by providing decision-making for complex scenarios as well as saving time and effort. The majority of these frameworks lack of parallel execution. DISSECT-CF is one of the frameworks that intro-

duced an improvement in performance of Infrastructure as a Service (IaaS) simulation. Although DISSECT-CF execution time is faster than the majority, it still executes sequentially. This paper introduces parallel execution to the most abstract subsystem in DISSECT-CF (event system). The new subsystem detects when multiple events occur at a specific time and then multi-threads these events. The number of independent frequent events, plays a crucial role to invoke the new subsystem and increase the performance. Achieving a high degree of repeated events leads to better performance. We focused on time management scenarios as a part of simulation to show the leverage of parallelism.

[C7] Al-Haboobi, A., & Kecskeméti, G. (2021). Improving Existing WMS for Reduced Makespan of Workflows with Lambda. In: Euro-Par 2020: Parallel Processing Workshops : Euro-Par 2020 International Workshops, Warsaw, Poland, August 24–25, 2020, Revised Selected Papers, Springer Nature Switzerland, 373 p. pp. 261-272. `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31925040`

Scientific workflows are increasingly important for complex scientific applications. Recently, Function as a Service (FaaS) has emerged as a platform for processing non-interactive tasks. FaaS (such as AWS Lambda and Google Cloud Functions) can play an important role in processing scientific workflows. DEWE v3 is one of the Workflow Management Systems (WMSs) that already had foundations for processing workflows with cloud functions. In this paper, we have modified the job dispatch algorithm of DEWE v3 on a function environment to reduce data dependency transfers. Our modified algorithm schedules jobs with precedence constraints to be executed in a single function invocation. Therefore, later jobs can utilise output files generated from their predecessor job in the same invocation, which can reduce the makespan of workflow execution. We have evaluated the improved scheduling algorithm and the original with small- and large-scale Montage workflows. The experimental results showed that our algorithm can reduce the overall makespan in contrast to the original DEWE v3 by about 10%.

[D1] Pflanzner, T., Mobile Simulation Environment for Investigating IoT-Cloud Systems. Doctoral thesis, Submitted to Doctoral School of Computer Science, October, 2020. Defended in January, 2021. `http://doktori.bibl.u-szeged.hu/10666/`

In the phase of the latest Internet evolution, a high number of powerful devices appeared on the network, which led to the birth of the IoT paradigm. As a result, cloud technology providers responded to this trend, and broadened their services to attract users of IoT applications. In this new world, IoT developers are facing many challenges: they need to purchase smart devices, to design and build-up a network of these components, to develop IoT applications, to test and evaluate these applications on the created system and finally, to fine-tune the applications based on the evaluation results. In this work we provided means to overcome some of these challenges. We developed a complex, semi-simulation environment composed

the following tools: an Android-based, mobile device simulator called MobIoTSim capable of simulating up to hundreds of IoT devices, a customizable cloud gateway service that is able to manage the simulated devices, and an IoT trace archive service called SUMMON, which can be used to gather real-world sensor data to be fed to the simulation experiments. We exemplified the usage of this complex simulation environment, and evaluated its device management scalability and responsiveness.

2.2 Scientific results in the 2nd year

In the second year, we continued our research by designing and implementing new, mobility-related IoT applications using fog and cloud systems, building on the results of the first year. We have also extended the DISSECT-CF-Fog simulator by adding new components (e.g. with the implemented actuator model). With the extended simulator, we modeled and analyzed the behaviour of several use cases, using real sensor data. We also analyzed and improved data management algorithms based on a fuzzy method. With the involvement of several BSc and MSc students, we are also continuing our planned, related research directions in blockchain integration, workflow support, and trace analysis.

We held our second annual meeting on November 23, 2021, where all project participants presented their research progress and results. In the project proposal we planned 2 international conference presentations and 1 journal article as scientific results for the second year. We managed to significantly overperform this plan by publishing 5 journal articles, 1 international conference papers and 2 book chapters, and submitted 1 doctoral thesis.

Research tasks in the proposal for the second year, their corresponding publications and the summary of the main results of the papers:

- Task 2.1 Designing and developing simulated IoT applications using the behaviour model: [J7], [C8], [C10], [J10]
- Task 2.2 Developing efficient and privacy-aware sensor data management algorithms using fuzzy methods: [C9], [B1], [B2]
- Task 2.3 Designing and developing fuzzy-based algorithms for Fog and Cloud service selection for executing the proposed IoT use cases: [J6], [C9]
- Task 2.4 Evaluating the proposed use cases and data management algorithms with the simulators in terms of scalability, execution efficiency, and privacy: [J7], [J8], [J9], [J11], [B1]

[J7] Markus A., Biro M., Kecskemeti G., Kertesz A., Actuator behaviour modelling in IoT-Fog-Cloud simulation. PEERJ COMPUTER SCIENCE, 7 Paper: e651 , 27 p.

(2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32204080

The inevitable evolution of information technology has led to the creation of IoT-Fog-Cloud systems, which can manage up to billions of smart devices, sensors and actuators connected through the Internet. These components continuously generate large amounts of data that can be stored and processed by Cloud and fog services. The investigation and detailed analysis of such complex system behaviour can be fostered by simulation solutions. The currently available, related simulation tools were lacking a generic actuator model including mobility management. In this paper, we presented the main results of our project, which is the extension of the DISSECT-CF-Fog simulator. It is able to support the analysis of arbitrary actuator events and mobility capabilities of IoT devices in IoT-Fog-Cloud systems. The main contributions of this work are: (i) a generic actuator model and its implementation in DISSECT-CF-Fog, and (ii) the evaluation of its use through logistics and healthcare scenarios. Our results showed that we can successfully model IoMT systems and behavioural changes of actuators in IoT-Fog-Cloud systems in general, and analyse their management issues in terms of usage cost and execution time.

[J8] Baniata H., Kertesz A., FoBSim: an extensible open-source simulation tool for integrated fog-blockchain systems. PEERJ COMPUTER SCIENCE, 7 Paper: e431 (2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31971886

The Blockchain (BC) technology was proven as to be an enhancement factor for security, decentralization, and reliability in distributed systems, leading to be successfully implemented in cryptocurrency industries. Fog computing (FC) is one of the recently emerged paradigms that needs to be improved to serve IoT environments of the future. As hundreds of projects, ideas, and systems were proposed, one can find a great R&D potential for integrating BC and FC technologies. To validate an integrated Fog-Blockchain protocol or method implementation, before the deployment phase, a suitable and accurate simulation environment is needed. Such validation should save a great deal of costs and efforts on researchers and companies adopting this integration. In this paper, we introduced FoBSim, a Fog-Blockchain simulator to ease the experimentation and validation of integrated Fog-Blockchain approaches. According to our proposed workflow of simulation, we implemented different Consensus Algorithms (CA), different deployment options of the BC in the FC architecture, and different functionalities of the BC in the simulation. Finally, we simulated several case studies analyzed the obtained results, which showed that deploying the BC network in the fog layer shows enhanced efficiency in terms of total run time and total storage cost.

[J9] Baniata H., Sami M., Kertesz A., Assessing anthropogenic heat flux of public cloud data centers: current and future trends. PEERJ COMPUTER SCIENCE, 7 Paper:

e478 , 18 p. (2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31999473

Global average temperature had been significantly increasing during the past century, mainly due to the growing rates of greenhouse gas (GHG) emissions, leading to a global warming problem. Many research works indicated other causes of this problem, such as the anthropogenic heat flux (AHF). Cloud computing (CC) data centers (DC), for example, perform massive computational tasks for end users, leading to emit huge amounts of waste heat towards the surrounding (local) atmosphere in the form of AHF. Out of the total power consumption of a public cloud DC, nearly 10% is wasted in the form of heat. In this paper, we quantitatively and qualitatively analyzed the current state of AHF emissions of the top three cloud service providers (i.e., Google, Azure and Amazon) according to their average energy consumption and the global distribution of their DCs. We found that Microsoft Azure emitted the highest amounts of AHF, followed by Amazon and Google, respectively. We also found that Europe is the most negatively affected by AHF of public DCs, due to its small area relative to other continents and the large number of cloud DCs within. We also provided future trends estimations of AHF densities. In one of the presented scenarios, our estimations predicted that by 2100, AHF of public clouds DCs will reach 0.01 Wm^{-2}.

[J10] Sallo, D. H., Kecskemeti G., A Parallel Event System for Large-Scale Cloud Simulations in DISSECT-CF. Acta Cybernetica, 25(2), 469-484. (2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32531834

Discrete Event Simulation (DES) frameworks gained significant popularity to support and evaluate cloud computing environments. They support decision-making for complex scenarios, saving time and effort. The majority of these frameworks lack parallel execution. In spite being a sequential framework, DISSECT-CF introduced significant performance improvements when simulating Infrastructure as a Service (IaaS) clouds. Even with these improvements over the state-of-the-art sequential simulators, there are several scenarios (e.g., large scale Internet of Things or serverless computing systems), which DISSECT-CF would not simulate in a timely fashion. In this paper, we introduced parallel execution to its most abstract subsystem of DISSECT-CF: the event system. The new event subsystem detects, when multiple events occur at a specific time instance of the simulation, and decides to execute them either on a parallel or a sequential fashion. This decision is mainly based on the number of independent events and the expected workload of a particular event. In our evaluation, we focused exclusively on time management scenarios. While we did so, we ensured the behaviour of the events should be equivalent to realistic, larger-scale simulation scenarios. This allowed us to understand the effects of parallelism on the whole framework, while we also shown the gains of the new system compared to the old sequential one. With regards to scaling, we observed it

to be proportional to the number of cores in the utilised host.

[J11] Al-Haboobi, A., Kecskemeti, G., Execution Time Reduction in Function Oriented Scientific Workflows. Acta Cybernetica, 25(2), 131-150. (2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32531775

Scientific workflows have been an increasingly important research area of distributed systems (such as CC). Researchers have shown an increased interest in the automated processing of scientific applications such as workflows. Recently, Function as a Service (FaaS) has emerged as a novel distributed systems platform for processing non-interactive applications. FaaS has limitations in resource use (e.g., CPU and RAM), as well as state management. Initial studies have already demonstrated using FaaS for processing scientific workflows. DEWE v3 is a tool that executes workflows in this fashion, but it often suffers from duplicate data transfers, while using FaaS. This behaviour is due to the handling of intermediate data dependencies after and before each function invocation. These data dependencies could fill the temporary storage of the function environment. In this paper we proposed a solution that altered the job dispatch algorithm of DEWE v3 to reduce data dependency transfers. Our developed algorithm schedules jobs with precedence requirements to primarily run in the same function invocation. We evaluated our proposed algorithm and the original algorithm with small- and large-scale Montage workflows. Our results showed that the improved system can reduce the total workflow execution time of scientific workflows over DEWE v3 by about 10% when using AWS Lambda.

[C8] Markus A., Dombi J. D., Kertesz A. (2021). Location-aware Task Allocation Strategies for IoT-Fog-Cloud Environments. In: Cristina, Ceballos (ed.) 2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), IEEE, pp. 185-192. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31989625

In this paper, we presented a revision and extension of the DISSECT-CF-Fog simulation environment, which provides a more detailed fog model by enhanced location awareness and multi-layer fog node management to foster IoT application execution. To validate our proposal, we developed different, basic task allocation strategies for IoT-Fog-Cloud environments, and also proposed a sophisticated fuzzy-based allocation algorithm using the Pliant system that is able to react to multiple system parameter changes. Finally, we evaluated these strategies for different scenarios in terms of cost, network utilisation, data management and application makespan.

[C9] Kertesz A., Baniata H. (2022). Consistency Analysis of Distributed Ledgers in Fog-Enhanced Blockchains. In: Euro-Par 2021: Parallel Processing Workshops : Euro-Par 2021 International Workshops, Lisbon, Portugal, August 30-31, 2021, Revised Selected Papers, Springer International Publishing, pp. 393-404. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32930145

Both revolutionary technologies of Fog Computing (FC) and Blockchain (BC) serve as enablers for enhanced, people-centric trusted applications. In this paper, we addressed the reliability of fog-enhanced BC systems by analyzing the forking phenomenon under different conditions, and provided a reliable Distributed Ledger (DL) consistency assessment. We used the FoBSim tool that is specifically designed to mimic and emulate realistic FC-BC integration, in which we deployed the Proof-of-Work (PoW) consensus algorithm and analyzed the forking probability under fluctuating conditions. Based on our results, we also proposed an inconsistency formula, which can quantitatively describe how consistent the DL in a BC system can be. Finally, we showed how to deploy this formula in a decision making model for indicating optimal deployment features of a BC network in a Fog-enhanced system.

[C10] Sallo, D. H., Kecskemeti G. (2022). Towards Generating Realistic Trace for Simulating Functions-as-a-Service. In: Euro-Par 2021: Parallel Processing Workshops: Euro-Par 2021 International Workshops, Lisbon, Portugal, August 30-31, 2021, Revised Selected Papers, Springer International Publishing, pp. 428-439. Paper: Chapter 34. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32875896

Serverless computing is a step forward to provide a cloud environment that responds to user requests by mainly focusing on managing infrastructure, resources and configurations. Despite the widespread use of cloud simulators, they are still mainly focused on supporting more traditional Infrastructure-as-a-Service scenarios and this reduces their applicability in the serverless and function as a service domains. Moreover, workload traces typically employed by IaaS simulators to represent user behaviour, are not well adoptable for serverless model. More realistic and serverless-like traces are essential to simulate and predict the behaviour of functions in serverless systems. In this paper, we focused on generating realistic traces for simulating serverless computing platforms. The generated traces produced by our approach are based on the Azure Functions dataset, and they are readily applicable in DISSECT-CF. We validated the generated approach using the coefficient of determination, which showed very good values for the average and percentiles of the execution time and memory. To demonstrate the benefits of the generated traces, we introduced a rudimentary model for serverless systems to DISSECT-CF. Our evaluation showed that our workloads were realistic and closely followed the behaviour of Azure's function as a service component.

[B1] Gültekin-Várkonyi G., Kertész A., Váradi Sz., Application of the General Data Protection Regulation for Social Robots in Smart Cities. In: Augusto, Juan Carlos (eds.) Handbook of Smart Cities. Springer International Publishing, 1, 25 p. (2021) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31827389

The recently emerged advances in ICT, such as cloud computing, the Internet of Things, and Artificial Intelligence, enabled the rapid creation of smart environments.

Smart devices have appeared in our everyday life gaining access to personal data by monitoring our behaviour and needs. To react to this new situation, the European Union introduced the General Data Protection Regulation (GDPR) in 2018 that must be applied as the general legal framework for personal data protection in smart environments as well. In this chapter, we provided discussions in the EU-related data protection acts in this new era, and we analyzed how smart appliances, specifically social robots, are used in AI technologies of smart homes. We also defined a case study of utilizing smart social robots in smart homes, and analyzed its privacy and data protection implications.

[B2] Márkus A., Kertész A., Investigating IoT Application Behaviour in Simulated Fog Environments. In: Helfert, Markus; Pahl, Claus; Ferguson, Donald (eds.) Cloud Computing and Services Science, Springer International Publishing, pp. 258-276. (2021) `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31933397`

Sensors and devices of IoT applications need big data to be stored, processed and analysed, and cloud systems offer suitable and scalable solutions for them. Recently fog nodes are utilized to provide data management functionalities closer to users with enhanced privacy and quality, giving birth to the creation of IoT-Fog-Cloud systems. Such infrastructures are so complex that they need simulators for planning, designing and analysis. Though cloud simulation already has a large number of literature, the simulation of fog systems is still evolving. In this paper, we investigated some of the current fog simulation approaches and compared two of them providing the broadest fog modeling features. We also performed evaluations of executing IoT applications in hybrid, Fog-Cloud architectures to show possible advantages of different setups matching different IoT behaviour.

2.3 Scientific results in the 3rd year

In the third year, we continued our research by finalizing the IoT workflow and actuator model based on the results of the second year. We adapted our actuator implementation to the field of the latest computing paradigm, Dew Computing, as part of an international collaboration, and evaluated it with two real-world mobility use cases. We have extended our fog and blockchain integration methods to enable data protection compliance and privacy protection with an optimized neighbor selection protocol.

We held our third annual meeting on November 15, 2022, where all participants presented their research progress and results. For the third year of the project, we planned 2 international conference presentations and 3 journal articles as scientific results. We managed to overperform this plan by publishing 3 journal articles, 3 international conference papers, and submitted 2 doctoral theses.

Research tasks in the proposal for the third year, their corresponding publications and the summary of the main results of the papers:

- Task 3.1 Designing and developing optimized algorithms for IoT data management based on the previous evaluations: [J11], [J12], [C9]
- Task 3.2 Evaluating the optimized IoT data management algorithms: [C10], [C11]
- Task 3.3 Developing a methodology for simulating IoT-Fog-Cloud systems exploiting IoT application behaviour modelling: [J7], [J13], [J14], [D2], [D3]

[J12] Baniata H., Anaqreh A., Kertesz A., DONS: Dynamic Optimized Neighbor Selection for smart blockchain networks. FUTURE GENERATION COMPUTER SYSTEMS, 130, pp. 75-90. (2022) `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;32588616`

Blockchain (BC) systems mainly depend on the consistent state of the Distributed Ledger (DL) at different logical and physical places of the network. The majority of network nodes need to be enforced to use one or both of the following approaches to remain consistent: (i) to wait for certain delays (i.e. by requesting a hard puzzle solution as in PoW) (ii) to propagate shared data through shortest possible paths within the network. The first approach may cause higher energy consumption and/or lower throughput rates if not optimized, and in many cases these features are conventionally fixed. Therefore, it is preferred to enhance the second approach with some optimization. Previous works for this approach may violate the identity privacy of miners, only locally optimize the Neighbor Selection method (NS), do not consider the dynamicity of the network, or require the nodes to know the precise size of the network at all times. In this paper we addressed these issues, and proposed a Dynamic and Optimized NS protocol called DONS. It uses a novel privacy-aware leader election within the public BC called AnoLE, where the leader anonymously solves the The Minimum Spanning Tree problem (MST) of the network in polynomial time. Consequently, miners are informed about the optimum NS according to the current state of network topology. We analytically evaluated the complexity, the security and the privacy of the proposed protocols against state-of-the-art MST solutions for DLs and well known attacks. Additionally, we experimentally showed that the proposed protocols outperform state-of-the-art NS solutions for public BCs.

[J13] Markus A., Biro M., Skala K., Sojat Z., Kertesz A., Modeling Dew Computing in DISSECT-CF-Fog. APPLIED SCIENCES-BASEL, 12:17, Paper: 8809, 12 p. (2022) `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;33079396`

IoT services need to be managed seamlessly to ensure adequate Quality of Service (QoS), due to the mobility of devices or the temporary periods without an internet connection. Such domains are combined under the auspices of Dew computing, as in critical cases, extending an IoT service to the end user's device is a feasible task. Such scenarios can hardly be investigated at a large scale due to the lack of dedicated simulation environments. In this paper, we presented an extension of the

DISSECT-CF-Fog simulator with a Dew computing model, to enable the simulation of IoT-Dew-Fog systems in a cost-effective manner. In particular, we focused on service migration options for mobile devices and cases with temporary internet access limitations. Finally, we performed measurements of real-world use cases with the extended simulator as an evaluation. Our simulation results showed that the proposed proactive strategy reduces the processing time of IoT data, exploiting an IoT-Dew-Fog environment.

[J14] Al-Haboobi, A., Kecskemeti, G., Developing a Workflow Management System Simulation for Capturing Internal IaaS behavioural knowledge. Journal of Grid Computing. 20 pages. Accepted on 25th November, 2022.

Conducting real experiments for large-scale workflows is challenging, because they are very expensive and time consuming. A simulation is an alternative approach to a real experiment that can help evaluating the performance of workflow management systems (WMS) and optimise workflow management techniques. Although, there are several workflow simulators available today, they are often user-oriented and treat the cloud as a black box. Unfortunately, this behaviour prevents the evaluation of the infrastructure level impact of the various decisions made by the WMSs. To address these issues, in this paper we proposed a WMS simulator (called DISSECT-CF-WMS) as an extension to DISSECT-CF. DISSECT-CF-WMS enables better energy awareness by allowing the study of schedulers for physical machines. It also supports dynamic provisioning to meet the resource needs of the workflow application while considering the provisioning delay of a VM in the cloud. We evaluated our simulation extension by analysing several workflow applications on a given infrastructure.

[C11] Markus A. (2022). Towards Modelling IoT Workflows. In: 13th Conference of PhD Students in Computer Science: Volume of Short Papers, pp. 25-29. `https://m2.mtmt.hu/gui2/?mode=browse&params=publication;33295842`

The cooperation of distributed computing and Internet of Things (IoT) paradigms created numerous research challenges. Modern applications do not only compute certain tasks, but they support various events that make human life more colourful. Considering workflows in the IoT domain, opposed to general scientific workflows, the focus is on the optimal execution of predefined sequences of various steps, however, computational tasks can also be contained. Such a sequence of steps usually consists of performing a service call, receiving a data packet in the form of a message sent by an IoT device, or executing a computational task on a virtual machine. The development and testing of such IoT workflows and their management systems in real life can be complicated due to high costs and access limitations, therefore simulation solutions should be preferred. In this paper, we discussed the current capabilities of scientific workflow simulation environments, defined the needs of IoT workflow modelling support, and made suggestions for their future realisation.

[D2] Markus A., DISSECT-CF-Fog: A Simulation Environment for Analysing the Cloud-to-Thing Continuum. Doctoral thesis, Submitted to Doctoral School of Computer Science, November, 2022. https://doktori.bibl.u-szeged.hu/id/eprint/11551/

IoT-Fog-Cloud systems require significant investments in terms of design, development and operation, therefore, the use of simulators for their investigation is inevitable. There are a large number of simulators addressing the analysis of parts of these systems, however, it is obvious that only a state-of-the-art simulator is capable of modelling complex architectures in a realistic way, which meets modern challenges.

In this thesis, we presented a detailed survey and taxonomy of various IoT, cloud, and fog simulators in order to determine the key requirements of a compact and well-defined IoT-Fog-Cloud simulator. We also showed an in-depth analysis and a comparison of two major fog simulators. We introduced the IoT and the pricing extension of DISSECT-CF, that realizes a multi-cloud environment with resource allocation strategies. Finally, we presented the DISSECT-CF-Fog simulator, which is able to model a multi-layered fog topology with energy measurement, task allocation algorithms and, mobility and actuator events.

[D3] Baniata, H., Integrating Blockchain and Fog Computing Technologies for Efficient Privacy-preserving Systems. Doctoral thesis, Submitted to Doctoral School of Computer Science, November, 2022. https://doktori.bibl.u-szeged.hu/id/eprint/11555/

Blockchain technology (BC) is a distributed ledger technology in the form of a distributed transactional database, secured by cryptography, and governed by a consensus mechanism. The integration of FC with BC can result in more efficient services, in terms of latency and privacy, mostly required by Internet of Things systems. In this thesis, we addressed some of these integration challenges. After a detailed and extended literature review of related simulation tools and integration approaches, we developed FoBSim, a novel simulation tool that allows for reliable and realistic FC-BC integration simulation. It facilitates the simulation of different consensus algorithms and different applications, and allows to deploy the BC at different layers of an FC-enabled cloud system, with the advantage of easy parameterization of simulation scenarios. Using this tool, we experimentally proved how integrating FC and BC can provide enhancement in terms of latency and cost. Additionally, we analyzed different factors affecting distributed ledger consistency and trust, which motivated the development of novel methods for quantifying the consistency and reliability of BCs. Using these methods, we could introduce a decision-making model resulting in better integration potential of FC and BC technologies. We also designed two novel protocols aiming to enhance FC-BC integrated applications with optimized neighbour selection, and two privacy-preserving solutions where BC is

exploited to enhance FC efficiency in terms of latency. Finally, we designed and developed PriFoB, a novel Blockchain-based Fog-enhanced global accreditation and credential verification system.

3 Additional remarks

3.1 Submitted papers

Besides the results and publications detailed before, we have 4 other journal articles submitted and being under review by the time of writing:

- Gavua E. K., & Kecskemeti G. Improving MapReduce Speculative Executions with Global Snapshots. Submitted to Elsevier Parallel Computing in June, 2022.

  In this paper we proposed a new speculative execution approach, which estimates task run times with consistent global snapshots and K-Means clustering. Task progress is captured consistently during data processing, where two categories of tasks (fast and poor-performing) are identified with K-Means. A silhouette score is applied to determine whether to transfer the slow tasks to available nodes or not. If the transfer is approved, the slow tasks are processed as backup tasks; otherwise they are left to process on their original nodes. Nevertheless, the fast tasks are always forwarded to the next phase (i.e., reduce phase or data output) after the clustering decision. We evaluated our approach on different data centre configurations based on Hadoop cluster requirements from the industry.

- Gavua E. K., & Kecskemeti G. Analysing Cloud Auto-Scaling Mechanisms with an Abstract State Machine Model. Submitted to the International Journal of Cloud Computing in August, 2022.

  In this work we proposed new ways of analysing and comparing auto-scalers designed on different clouds. We defined a model of five ASM transition rules to reflect typical job execution phases, and presented comparisons of auto-scalers offered alongside the DISSECT-CF cloud simulator with our transition rules. We also validated and verified our model with test cases by applying a Computational Tree Logic (CTL) formulae using the CoreASM Model Checking Tool. These results proved that our model can provide insight into the behaviour of the auto-scalers, even without conducting real-life or simulated experiments.

- Sallo D. H., & Kecskemeti G. Enriching Computing Simulators by Generating Realistic Serverless Traces. Submitted to Journal of Cloud Computing in August, 2022.

  In this paper, we proposed an approach for generating realistic serverless traces. This new, genetic algorithm based approach improves the statistical properties of the generated traces. We also enabled arbitrary scaling of the workload defined in the traces, while maintaining real users' behaviour. These advances further support reproducibility in the serverless research community. We validated the results of our generator approach using the coefficient of determination, which showed that our generated workload closely matches the original dataset's characteristics, in

terms of execution time, memory utilisation as well as user participation percentage.

- Markus A., Al-Haboobi A., Kecskemeti G., Kertesz, A. Simulating IoT Workflows in DISSECT-CF-Fog. Submitted to Sensors in December, 2022.

  The modelling of IoT applications utilising resources of Cloud and Fog Computing is not straightforward, because they have to support various trigger-based events that make human life easier. The sequence of tasks, such as performing a service call, receiving a data packet in the form of a message sent by an IoT device, and managing actuators or executing a computational task on a virtual machine, are often associated with and composed to IoT workflows. The development and deployment of such IoT workflows and their management systems including communications and network operations in real life can be complicated due to high operation costs and access limitations, therefore simulation solutions often applied for such purposes. In this paper, we introduce a novel simulator extension of the DISSECT-CF-Fog simulator that leverages the workflow scheduling and its execution capabilities to model real-life IoT use-cases. We also show that state-of-the-art simulators typically omit the IoT factor in the case of the scientific workflow evaluation, therefore we present a scalability study focusing on scientific workflows, and the interoperability in DISSECT-CF-Fog between scientific and IoT workflows.

3.2 Dissemination

Regarding dissemination, in the past years we gave two invited talks at international events related to the project topics:

- A. Kertesz, Simulation and behaviour analysis of complex systems, CERCIRAS Training School 2022 on Time Predictability, Energy Budgeting, Simulation-based Analysis of complex systems, and Mixed-Criticality Systems, 2022.
- A. Kertesz, Integrating Fog Computing and Blockchain Technology for Applications with Enhanced Trust and Privacy. LSDVE'21 workshop of EuroPar'21, Lisbon, Portugal, 31 August, 2021.

We also presented our results at several project meetings within the framework of the CERCIRAS EU Cost Action, and promoted our DISSECT-CF-Fog simulator in the HiPEAC magazine:

- A. Kertesz, A. Markus, Modeling IoT-Fog-Cloud systems utilizing DISSECT-CF-Fog simulator. HiPEAC 67 magazine, 2022. `https://issuu.com/hipeac/docs/hipeacinfo_67_high_res/24`

3.3 Note on the project participants

Our originally submitted project proposal included unnamed researchers to be finalized upon the start of the project. Right after the start of the project we performed these allocations, and on December 19, 2019, we sent the names of the allocated participants and their corresponding tasks to the NKFI Office, which was approved on January 7, 2020. As a result we allocated the following additional researchers to the project, while the originally named researchers were not changed:

- Ali Sadeq Abdulhameed Al-Haboobi and Dilshad Hassan Sallo (University of Miskolc); Hamza Bassam Bani Ata, Dániel József Dombi and Dr. Szilvia Váradi (University of Szeged);