Event Based Systems in Iaşi

Project Number:	692178
Project Title:	EBSIS - Event Based Systems in Iaşi A Twinning between Universitatea Alexandru Ioan Cuza din Iaşi, Université de Neuchâtel and Technische Universität Dresden
Project Call Topic:	H2020-TWINN-2015 (Twinning)
Deliverable Number:	D1.3
Title of Deliverable:	Final rolling report on primary training actions
Date of Delivery:	2018.12.19

Executive Summary

This report includes details on the block lectures taught at UAIC by partner institutions during the third year of the project, as well as projected additions to the UAIC curriculum. The report also presents details on the master co-supervision actions organized between UAIC and the other project partners that started in the final year of the project.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 692178.

List of Contributors

Name	Organization
Emanuel Onica	UAIC
Diana Trandabăţ	UAIC
Hugues Mercier	UNINE
Rafael Pires	UNINE
Pascal Felber	UNINE
Etienne Rivière	UCL
André Martin	TUD
Martin Nowack	TUD
Christof Fetzer	TUD

Document Approval

	Name	Date
Approved by WP Leader	Christof Fetzer	2018.12.17
Approved by MB Member 1	Emanuel Onica	2018.12.17
Approved by MB Member 2	Pascal Felber	2018.12.17

Contents

List of Figures

1. Block lectures

In the third year of the project, the block lectures organized at the site of UAIC continued with five more sessions associated with tasks T1.1, T1.2 and T1.3 in WP1. As specified in the project work description, one or more lectures were organized for each of the task specifics – scalability, dependability and security of event based systems.

A first lecture associated with task T1.1 discussed multicore programming, and in particular techniques that are widely used nowadays in achieving vertical scalability of computer applications. Following the interest observed in the second year of the project, a second block lecture session was organized again on the topics of data compression. The topic as described in D1.2 lies at the intersection of all the three areas of interest of the project, but which we consider as part of T1.1 due to wider applications in increasing scalability and performance of distributed systems. The two lectures corresponding to task T1.1 were tutored by a staff member of TUD for the multicore programming lecture, and a staff member of UNINE for the data compression lecture, respectively.

The lecture corresponding to task T1.2 consisted in a third instantiation of a fundamental course on event based systems dependability that was highly appreciated by the previous two series of master students, and which was taught again by a staff member of TUD.

The lecture organized as part of task T1.3 targeted the emerging area of blockchain technologies in the context of their distributed nature and security traits. The lecture was taught by a professor of UCL. The planned involvement in the tasks considered UCL for task T1.1 and TUD for task T1.3. However, due to the particular cross-domain expertise of tutors with respect to the two lectures, the consortium preferred switching the partners contribution between the two corresponding tasks.

A fifth lecture was organized at the site of UNINE, being taught by an associate professor of UAIC. This lecture was initially planned as part of task T1.3, which is focused on event based systems security. However, the partners commonly agreed that it would be in the best interest of the project to organize a lecture oriented on natural language processing (NLP) instead. Collaboration on future security research topics was already established between partners. The chosen topic of NLP was meant to bring the information retrieval research groups in the two institutions closer to each other. This is a research direction that has strong ties with event based processing because multiple emerging systems rely on intelligent retrieval of information from human communication such as data streams of tweets.

As before, all the organized lectures were integrated within the master studies curriculum of UAIC and UNINE, respectively, but the attendance included also senior bachelor students who were interested in the discussed topics. The organization of the lectures was similar with prior years; condensed sessions of not more than two days were planned to allow integrating these within the academic schedule of the students.

A follow-up of the high student interest observed for the blockchain technologies lecture was the consideration of introducing a permanent lecture covering the distributed and security aspects of this topic in the curriculum of UAIC. This is currently projected for the following academic year. As another follow-up to the lectures, three master students expressed interest in pursuing research topics in the area of the project. Consequently, new master theses co-supervisions were started as part of task T1.4, which we further discuss in Section 2. In the following, we provide a more detailed summary of the topics presented in each of the block lectures.

1.1 Block lectures focused on event based systems scalability (T1.1)

Multicore Programming
Martin Nowack (TU Dresden)
Dates: March 21-22, 2018

Lecture description:

Part 1: Introduction to Multicore Programming Using Lock-Free Data Structures

This part introduced multicore programming and started with the sequential implementation of a simple data structure. The lecture demonstrated how this can be transformed to a concurrent but simple lock-based implementation and finally to a lock-free implementation. The major goal was to get a first understanding of those three different approaches and how to continue from there.

Part 2: Performance Profiling and Evaluation of Singlecore and Multicore Applications

For a successful utilisation of multiple cores within an application, we need to understand its performance behaviour in order to improve it. This lecture introduced and provided hands-on experiments for testing and profiling performance in order to direct our optimisation work. "Go" was used as an example language. The description focused both on single and multicore implementations.

Figure 1 *Multicore Programming* lecture session

Data Compression from Shannon to Google

Dr. Hugues Mercier (Université de Neuchâtel)

Dates: May 15-16, 2018

As described in the introduction section, following the high interest observed during the previous project year, the data compression lectures were repeated for the new series of master students enrolled in the 2017-2018 academic year. The lectures followed a similar structure, but included some content updates, especially on the recently adopted standards leveraging asymmetric numeral systems.

Lecture description:

Starting from the original work of Claude Shannon, the lectures introduced the general concepts of data compression, discussed the tradeoffs between theoretical guarantees and practical implementations, and presented the 2014 breakthrough that led, after 30 years of stagnation, to an explosion of new codes used by Facebook, Apple and Google.

The lecture sessions covered the following topics:

- Introduction to entropy, probability and Markov chains
- Introduction to data compression
- Shannon source coding theorem
- Huffman coding
- Dictionary techniques
- Arithmetic coding
- Asymmetric numerical systems and modern standards

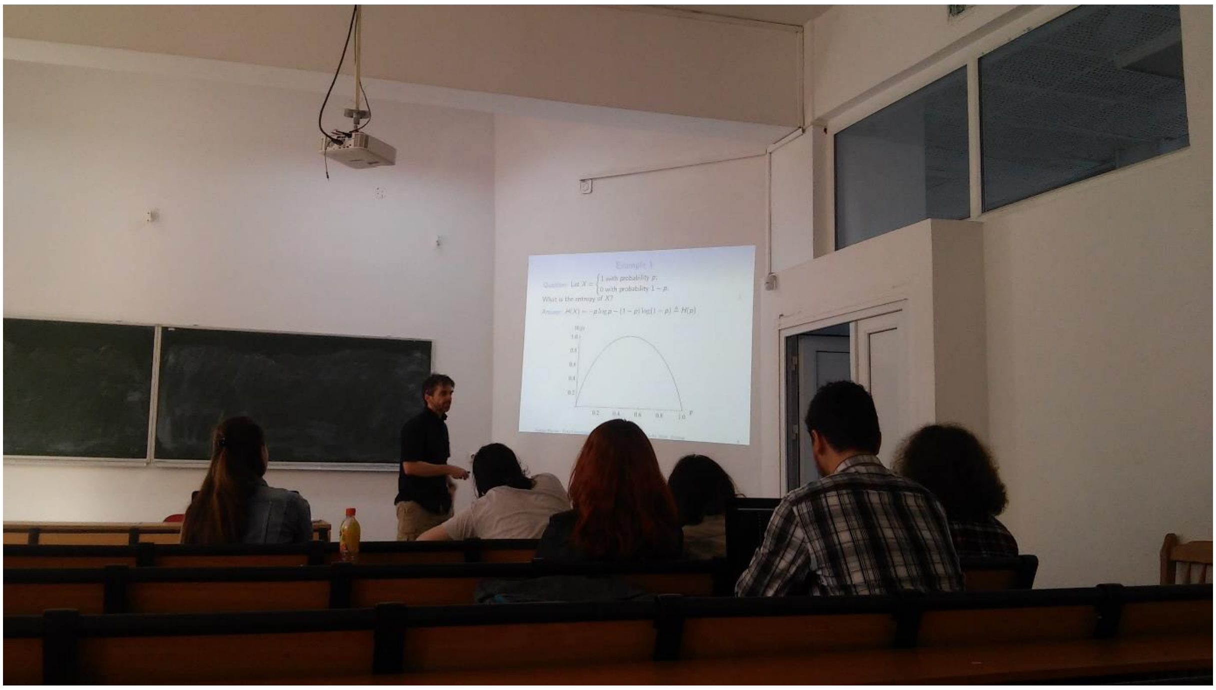

Figure 2 *Data Compression* lecture session

1.2 Block lectures focused on event based systems dependability (T1.2)

Introduction to Event Based Systems Dependability

Dr. André Martin (TU Dresden)

Dates: May 3-4, 2018

This block lecture followed the same structure as in the previous two years, being addressed to the new series of master students enrolled in the 2018-2019 academic year. Besides the good feedback received after the lecture sessions in the previous years, one important reason of repeating the same topics lies in the fact that the lecture content is fundamental for the area of event-based systems.

Lecture description:

Introduction to Event Stream Processing/Dependability

The first lecture of the dependability course introduced students to the fundamentals of Event Stream Processing (ESP) systems such as the data, query and execution model. To get familiar with ESP systems, several real-world streaming applications and their implementations originating from the annual DEBS challenge were discussed. Two of the applications presented covered the following use cases: energy consumption prediction and an online analysis of taxi rides in the NYC area. An overview of dependability fundamentals such as common failures models, recovery guarantees etc. tailored to ESP systems was given. The lecture concluded with an overview of common fault tolerance mechanisms used in ESP systems.

Passive Replication

The second lecture of the dependability class took a closer look at recovery protocols used in passive replication. This was followed by challenges with passive replication such as the overhead introduced by check-pointing and event ordering. Several approaches based on latest research that allow to reduce the overhead of check-pointing and event ordering were presented.

Active Replication

The third lecture in this course was dedicated to active replication, an alternative to passive replication. The lecture covered the principles of active replication and challenges such as how to achieve consistency across replicas. Several different protocols to achieve consistency using a deterministic merge were presented. Finally, the lecture concluded with a discussion on how the overhead can be reduced for several classes of applications using slightly weaker consistency semantics.

Adaptive Fault Tolerance

The last lecture of the course covered an adaptive approach where several replication strategies are combined and adapted during run-time. This included two different approaches tailored to cloud environments that allow reducing the resource overhead of active replication and improve availability at the same time.

1.3 Block lectures focused on event based systems security (T1.3)

Bitcoin, the Blockchain and Beyond
Prof. Etienne Rivière (UC Louvain)
Date: May 9, 2018

Lecture description:

The lecture covered the design and implementation of innovative and trustworthy applications using blockchain technologies.

First, it discussed the famous Bitcoin cryptocurrency and defined how this allows massive-scale distributed consensus in the presence of selfish and malicious participants. Details were presented on how transactions, blocks, and the blockchain are implemented in Bitcoin. Following, the lecture included a discussion on alternative implementations of these building blocks and of blockchains in general, and how they enable novel decentralized applications and trust models.

Finally, the lecture covered applications of smart contracts, efficient consensus protocols, and permissioned blockchains.

Figure 3 *Bitcoin, the Blockchain and Beyond* lecture session

Introduction to Natural Language Processing

Assoc. Prof. Diana Trandabăţ (Alexandru Ioan Cuza University of Iaşi)

Date: September 26, 2018

As discussed in the section introduction, it was preferred by the consortium partners that the session of security oriented lectures originally planned, to be replaced by a lecture in the area of information retrieval. There exist multiple applications of natural language processing in information security. Although the lecture did not specifically target these, the fundamental notions that were presented can be used in this field. The consortium already organized multiple activities targeting security topics, therefore it was of more benefit to target a complementary topic. The activity also created context for meetings and discussions with the information retrieval group at UNINE, where the lecture was organized, which increased the potential of future joint research. This is in line with the project objectives.

Lecture description:

Artificial Intelligence enhances computers with human-like appearance, behavior and even judgment. In an attempt to learn how people communicate, computers use language models derived from human knowledge. While still far from completely understanding hidden messages in political discourses, computer and social scientists have joined efforts in modeling a human-like linguistic behavior. This is the field of Natural Language Processing (NLP), which was introduced during this block lecture. NLP's main focus is on language, either written or spoken. Texts live around us just as we live around them. At any instant, there are texts that people write, share, use to get informed, etc.

Research in NLP has resulted in massive amounts of structured linguistic data, such as ontologies, annotated corpora, various resources and tools. However, much of the information floating around us is unstructured, unannotated, raw data. Here is where the Natural Language Processing field intervenes, intending to analyze, structure, extract and generate meaning from the available sources of information, with an emphasis on real-life applications.

2. Co-supervision of master theses

The consortium partners proposed a variety of topics for attracting students towards research in the area of the project and starting a master thesis in such a direction. Three students manifested interest in the project topics during the 3rd year of the project and corresponding new master theses co-supervisions were started. We detail the topics that were chosen by the students, as they were proposed, in the following:

- *Improvement of the dissemination layer of cryptocurrencies using novel dissemination algorithms* (proposed by UNINE). Every cryptocurrency and distributed ledger based on blockchains has a communication layer used to disseminate the transactions to all the members of the network. This communication layer typically uses gossip-based algorithms that propagate information in a random fashion. Our initial analysis is that this communication layer is rather rudimentary and suffers from two main weaknesses. First, the tail of the distribution is rather fat, in the sense that it takes a long time to propagate the transactions to a small but important subset of nodes. Second, the diffusion of the transactions, while random, is not uniform. These two features lead to important security issues such as increased probability of double spending, peer isolation, competition between mining pools, and forking. They also degrade the performance of this layer by requiring a uselessly large number of messages between the peers to disseminate the transactions.

  The tasks of this thesis topic are:
  - Study the implementation of the dissemination layers of popular distributed ledgers (e.g., Bitcoin, Ethereum or Hyperledger);
  - Modify this dissemination layer using the more efficient dissemination algorithms that were designed by UNINE; these algorithms already exist, and while the student needs to understand them, they essentially need to be implemented and tested in a new setting;
  - Test the performance and security of the resulting changes using simulations and, even better, real-world historical traces.

  Attacks on Trusted Execution Environments (proposed by UNINE). With the emerging of cloud computing, security concerns have grown in importance. Since sensitive code and data may be transferred to potentially hostile environments (e.g., a multi-tenant third party infrastructure provider), measures must be taken to ensure systems' dependability. To that end, trusted execution environments (TEEs) have been put into focus, particularly in server-scale commodity hardware, such as Intel software guard extensions (SGX). Software security, however, is not a feature that may come out-of-the-box within the physical platform where it runs. It heavily depends on the software design, the size of its trusted computing base (TCB) and potentially unnoticed vulnerabilities that may come within production-ready software releases. With the purpose of assessing such systems, we propose as main task of elaborating the thesis, the development of a set of attacks (e.g., control-flow hijacking, cache and side-channel attacks) that could detect and allow further action to prevent such security threats.

- *Blockchain benchmarking* (proposed by UCL). The context of blockchains has gained a very strong momentum following the success of the Bitcoin cryptocurrency. A blockchain is an immutable, append-only database that is shared and agreed upon by all participants of a decentralized large-scale system, which do not necessarily mutually trust each other. At the hearth of a blockchain lies a byzantine fault-tolerant consensus protocol, whereby decentralized entities agree on the same content and evolution of the chain with very high probability. The original consensus protocol of Bitcoin, often called the Nakamoto consensus after the (pseudonymous) Bitcoin author(s), leverage a proof-of-work that requires validators of new blocks, called miners, to solve a computationally intensive puzzle to gain the right to publish a block and gain associated rewards. The most viable economic strategy remains to play by the rules of the protocol, making it resilient to byzantine behaviors unless an adversary controls a significant portion of the network. The Ethereum blockchain extends on the Bitcoin principle by allowing the execution of arbitrary code on top of the blockchain content in the form of smart contracts. Ethereum uses proof-of-work as its consensus protocol. Many proposals for novel fully decentralized consensus protocols have been made, including proof-of-stake, proof-of-authority, or, more recently, proof of space-time. At the same time, several projects target more controlled environments (e.g. cross-enterprise settings) and the permissioned blockchain model. In this model, a set of known servers act as members of a byzantine fault tolerant consensus group, stepping back from the full decentralization of Ethereum and similar but allowing better performance. The prevalent project in this space is HyperLedger.
  It is not always easy for a developer of a novel blockchain-based application to decide which blockchain or which blockchain configuration is, or is not appropriate for her project. Understanding the performance implications of the many complicated distributed systems design choices made in the implementation of a given blockchain is often beyond the competences of high-level application designers.

  The goal of this thesis is to make a step towards a better understandability of the performance implication of choosing one blockchain implementation or another. There are several possibilities open for the student, to be decided in agreement with the supervisors at the beginning of the project. These include for instance:
  - Building a benchmarking tool with a collection of smart contracts and associated KPI (key performance indicators) that can be ran on different blockchains in order to compare performance;
  - Building a simplified, modular, blockchain implementation that can be used as a canvas for studying the impact of certain design choices on performance, delays, throughput etc.; one of the possible uses of such a framework would be for teaching and education purposes, in order to give a practical idea of performance implications of blockchain design choices to application developers.

All the three students traveled to the partner sites for extended visits during which the work on the topics was started and initial guidelines were set. Regular periodical meetings were organized after their return and their progress is monitored by the co-supervisors. A promising advancement was registered by the date of writing of this deliverable in all three cases. In particular, for the 2nd topic –

Attacks on TEEs - a research paper including a significant contribution of the student is already in preparation to be submitted to a top conference. Due to their late starting date, the finalization of co-supervision and defense of all the three master theses is expected to happen after the end of the EBSIS project. However, the consortium partners involved will remain committed to a successful completion of the co-supervisions, relying on self-funds for this purpose.

The second series of co-supervisions consisting in two master theses oriented on anonymity topics that started during the 2nd year of the project, unfortunately did not reach a conclusion. The two students interrupted their activity towards pursuing a master degree on the respective chosen topics due to personal reasons. However, as a partial confirmation of the master theses co-supervision task (T1.4) role in the project scope, a second student involved in the first series of co-supervisions pursued research activity after his defense, co-authoring an article submitted to a conference[1], although its acceptance is pending.

[1] A first student in this series of co-supervisions was a co-author of a research paper while the co-supervision was ongoing, before his master thesis defense, as described in D1.2.