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:	D1.1
Title of Deliverable:	First rolling report on primary training actions
Date of Delivery:	2016.12.21

Executive Summary

This report includes details on the block lectures taught at UAIC by partner institutions during the first year of the project. The report covers the impact of these actions such as the additions to the teaching curriculum at UAIC and other related aspects.

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
Pascal Felber	UNINE
Etienne Rivière	UNINE
Alberto Ravagnani	UNINE
Hugues Mercier	UNINE
André Martin	TUD
Elke Franz	TUD
Christof Fetzer	TUD

Document Approval

	Name	Date
Approved by WP Leader	Christof Fetzer	2016.12.19
Approved by MB Member 1	Emanuel Onica	2016.12.19
Approved by MB Member 2	Pascal Felber	2016.12.19

Contents

List of Figures

1. Block lectures

Four series of block lectures were organized during the first year in the premises of UAIC, the coordinating institution, as part of the first work package WP1 in the EBSIS project. These lectures were integrated within the master studies curriculum. For efficiency and due to restrictions caused by the students' schedule, the block lectures were organized in a condensed manner in sessions spanning two days. Two staff members of UNINE and two staff members of TUD were involved as tutoring personnel in this activity. For optimizing costs, each lecture was organized within the same period of time as the scientific seminars given by the staff members involved also in the corresponding project tasks included in WP2. Each lecture addressed a topic within the scientific scope of the project or in the correlated areas of interest. The attendance was formed mostly of students enrolled in the master programme of the Faculty of Computer Science in UAIC, but were also advertised to senior bachelor students and academic staff. The main purpose was to attract students to research topics in relation to the project, which could be further pursued as part of the elaboration of their dissertation thesis. We provide below a synthetic overview of each lecture that was organized and the correspondence with the specific tasks within the project work package.

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

Introduction to Concurrency
Prof. Pascal Felber (Université de Neuchâtel)
Dates: May 18-19, 2016

The lectures were integrated with the curriculum of the Distributed Systems master programme, but were attended also by students enrolled in the Systems Engineering and Computational Optimization master programmes. The course covered foundations of concurrent systems and multiprocessor synchronization. Students had the opportunity to learn about the basics of multicore programming, a new paradigm in computer science. The course included fundamental notions related to concurrent objects and practical algorithms that improve scalability of multithreaded event based systems applications. Some topics that were addressed are the following:

- Spin Locks and Contention Management
- Linked Lists: Locking, Lock-Free, and Beyond
- Data Parallelism
- Transactional Memory (Software and Hardware)

Figure 1 ***Introduction to Concurrency*** **lecture session**

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

Introduction to Event Based Systems Dependability
Dr. André Martin (TU Dresden)
Dates: May 12-13, 2016

The lectures were integrated with the curriculum of the Distributed Systems master programme. Also, students enrolled in the Information Security masters attended. The contents were structured in four sections:

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. In order to get familiar with ESP systems, several real world streaming applications and their implementations originating from the annual DEBS challenge were discussed. Two applications presented covered the following use cases: energy consumption prediction and an online analysis of taxi rides in the NYC area. An overview about dependability fundamentals such as common failures models, recovery guarantees etc. tailored to ESP systems was given. The lecture concluded with an overview about 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 were presented that allow to reduce the overhead of check-pointing and event ordering.

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.

Figure 2 ***Introduction to Event Based Systems Dependability*** **lecture session**

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

Two lectures were dedicated to the area of network coding and were jointly taught by UNINE and TUD being integrated with the curriculum of the Information Security master programme. Although not directly covering event-based systems, the topics presented can have multiple applications in this area, both in security and as well in data transfer optimization. We provide in the following the description of the two lectures.

Introduction to Linear Network Coding
Alberto Ravagnani (Université de Neuchâtel)
Date: April 25-26, 2016

In 2000, it was discovered that the information rate of a network communication might be improved by employing coding at the intermediate nodes of the network. If a source of information attempts to transmit messages to certain receivers via a network, then a network coding strategy would consist in injecting the messages in the network, and then make the intermediate nodes cooperate to spread information faster towards the receivers.

It was shown that a multicasting technique allows in practice to increase the number of delivered messages per channel use. This phenomenon was described in an example entitled the "Butterfly network", which was discussed in detail as part of the lecture.

In 2002 it was shown that the maximum multicast rate of a network communication is bounded by a certain graph-theoretic invariant associated to the underlying network. The invariant is well-known in mathematics as the min-cut between vertices (viewed as a graph).

It was shown that the min-cut bound can always be achieved letting the intermediate nodes of the network perform appropriate linear operations on the received inputs, provided that the messages are vectors over a sufficiently large field. The result is known as the "Max-Flow-Min-Cut" theorem, and it essentially gave birth to linear network coding as a research field. The mini-course presented a partial proof for the theorem that uses an original algebraic idea based on polynomials.

The "Max-Flow-Min-Cut" theorem shows the existence of linear node operations that achieve the maximum rate, but it does not provide an effective method to concretely find them. A second question investigated in this introductory course was therefore *How can one concretely design node operations that achieve the maximum multicast rate over a given network?*

It was shown in the research that the maximum multicast rate over any network is achieved with high probability letting the intermediate nodes perform random linear operations on the received inputs. Motivated by these results, a mathematical description of network communications in which the transmitted messages are vector spaces rather than vectors was proposed. This model was based on a new class of error-correcting codes called subspace codes, which were presented in the last part of the lecture.

A summary of the topics covered and briefly described above is the following:

- Motivating examples: the "Butterfly network". Routing versus coding.
- Definition of network and mincut between two vertices. Examples.
- The min-cut bound (idea of the proof).
- The "Max-Flow-Min-Cut" theorem (with proof).
- Subspace codes and subspace error correction.

Figure 3 ***Introduction to Linear Network Coding*** **lecture session**

Introduction to Secure Network Coding
Elke Franz (TU Dresden)
Dates: April 27-28, 2016

Network Coding is a promising approach for increasing throughput, energy efficiency, and robustness of data transmission. However, mere network coding is vulnerable to passive as well as active attacks. Even if network coding already provides a certain level of security against eavesdropping (passive attacks), this inherent security is not sufficient against a stronger attacker who may observe more links or even intermediate nodes in the network. Moreover, network coding is especially vulnerable to pollution attacks (active attacks). Even one single packet can influence the whole subsequent downstream processing of the packets.

Because of the vulnerability of network coding to attacks, there has been a lot of research into security of network coding in the past few years. The approaches discussed in this short course were based on Practical Network Coding (PNC). PNC provides a decentralized implementation of Random Linear Network Coding. All receiving nodes are able to decode the combined data packets they receive without the knowledge of coefficients randomly chosen by forwarding nodes for the computation of linear combinations.

Confidentiality of transmitted data can also be enforced by means of end-to-end encryption of the data before network coding is applied. However, network coding schemes providing confidentiality are more efficient than end-to-end encryption.

Due to the damaging influence of pollution attacks, the majority of the suggested secure network coding schemes deal with that threat. Known approaches are based on information theory, network error correction, or cryptography. The course especially focused on the latter. Common cryptographic approaches for authentication like digital signatures or Message Authentication Codes (MACs) cannot be directly applied to network coding since the packets are modified by forwarding nodes. Consequently, MACs as well as digital signatures would become invalid after one hop. To overcome this problem, authenticated checksums, homomorphic hashes, homomorphic MACs, and homomorphic signatures have been suggested for the detection of pollution attacks in network coding.

A synthetic summary of the topics described above and covered by the lecture is as following:

- Practical Network Coding
- Security Requirements for Network Coding
- Excursus: Cryptography
- Network Coding Schemes Providing Confidentiality
- Detection of Pollution Attacks

2. Other aspects related to first work package activities (primary training actions)

The activities covered by task T1.4 *Co-supervision of master theses* were also started during the first project year. Following the organized lectures three master students have shown interest in these topics proposed by UNINE and TUD partners that relate with the curriculum content. Two students have chosen topics proposed by Prof. Pascal Felber (*"Secure processing with SGX and map/reduce")* and Dr. Hugues Mercier (*"Reliable data storage using erasure correcting codes in distributed systems"*) from UniNE. One student has chosen a topic proposed by Dr. André Martin (*"Storm wrapper for fault-tolerant topology processing with Streamine3G"*) from TUD. All UAIC students started their activity during the Fall semester of the 2016 academic with an initial visit to partner institutions for discussions and guidance with their external co-supervisors. It is expected that the activity of students covered in their theses to possibly successfully address new open research problems.

Following the start of the project two new entire courses related to the project topic were introduced in the curriculum of the host institution. The *"Distributed Event Based Systems"* course was introduced as part of the Distributed Systems master programme, being explicitly linked with the EBSIS project. The block lectures included in T1.1 and T1.2 were actually integrated within the topics covered by this course during the Spring semester of the academic year 2016. The course was also integrated as an optional discipline in the curriculum of the Systems Engineering and Computational Optimization master programmes.
A second course, *"Practical aspects in Cloud Services Engineering"* was also introduced in the curriculum of the Systems Engineering master programme, in collaboration with local industry (Amazon Development Center). Although not related in a direct formal way with the EBSIS project, the course is highly related with the EBSIS technical topics showing the impact and growth of interest related to this area in the host institution, as well as increasing collaboration with external partners from industry.