DSIE Doctoral Symposium in informatics Engineering | 14

Proceedings

of the
9th Doctoral Symposium in Informatics Engineering

January 30th and 31st, 2014
Porto, Portugal

Editors:
A. Augusto Sousa
Eugénio Oliveira

www.fe.up.pt/dsie14

Sponsors:

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO
DEI Departamento de Engenharia Informática
CESCE SI SIA Group
ORBCOM IT & CONSULTING SERVICES
ATM SIS
VISUAL COMUNICAÇÃO
ITEN CPC IS – PROLÓGICA The new way
EPE Escola Profissional de Esposende
ZENDENSINO cooperativa de ensino IPRL
tadinense artes gráficas

COPYRIGHT

1ª Edição/ 1st Edition 2014
Editors: A. Augusto Sousa and Eugénio Oliveira
Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, 4200-465 Porto

DSIE'14 SECRETARIAT:
Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, s/n
4200-465 Porto, Portugal
Telephone: +351 22 508 21 34
Fax: +351 22 508 14 43

E-mail: dsie14@fe.up.pt
Symposium Website: http://www.fe.up.pt/dsie14

FOREWORD

STEERING COMMITTEE

These are difficult times. PhD students are engaged in an enterprise they often do not know whether the reward is at reach or not. However, these are also times for strong wills guiding their way ahead. PhD students organizing and collaborating with this event deserve our admiration for their efforts to keep alive their quest for new knowledge and better intellectual and technical skills.

2014 Doctoral Symposium in Informatics Engineering - DSIE'14 shapes the 9th edition of a scientific meeting usually organized by PhD students of the FEUP Doctoral Program in Informatics Engineering (ProDEI).

DSIE meetings have been held since the school year 2005/06 and the main goal has always been to provide a forum for discussion on, and demonstration of, the practical application of a variety of scientific research issues, particularly in the context of information technology, computer science and computer engineering. DSIE symposium comes out as a natural conclusion of mandatory ProDEI course called "Methodologies for Scientific Research" (MSR) leading to a formal evaluation of the students learned competencies.

The aim of this specific course (MSR) is to give students the opportunity to learn the processes, methodologies and best practices related to scientific research, particularly in the referred areas, as well as to improve their own capability to produce adequate scientific texts. With a mixed format based on multidisciplinary seminars and tutorials, the course culminates with the realization of DSIE meeting, seen as a kind of laboratory test for the concepts learned by students. In the scope of DSIE, students are expected to play various roles, such as authors of the articles, members of both scientific and organization committees, as well as reviewers, duly guided by more senior lecturers and professors.

DSIE event is then seen as a "leitmotif" for the students to be exposed to all facets of a scientific meeting associated with outstanding research activities in the area. Although still at an embryonic stage, and despite some of the papers still lack of maturity, we already can find some interesting research work or interesting perspectives about future work. At this time, it was not essential, nor even possible, for most of the students in the first semester of their PhD work, to produce strong and deep research results. However, we hope that the basic requirements for publishing an acceptable scientific paper have been fulfilled.

DSIE'14 Proceedings include 13 articles accepted in the previously defined context. They cover a large spectrum of topics in informatics / computer science and engineering areas and can be grouped according to four main clusters. These clusters include some different, although related topics, since, as it was expected, the paper themes are of a large diversity. These clusters are named as follows: Intelligent Systems and Machine learning (4 papers), Information Systems & Communication Networks (3 papers), Software Engineering & Security Systems (3 papers), Simulation Systems & High-Performance Computing (3 papers).

The complete DSIE'14 meeting encompasses a one and half a day program that includes also two invited talks by an outstanding researcher in Machine Learning and a successful entrepreneur in high tech knowledge transfer.

Professors responsible for ProDEI program current edition, are proud to participate in DSIE'14 meeting and would like to acknowledge all the students who have been deeply involved in the success of this event that, hopefully, will contribute for a better understanding of the themes that have been addressed during the above referred course, the best scientific research methods and the good practices for writing scientific papers and conveying novel ideas.

Eugénio Oliveira and Augusto Sousa (ProDEI)
January 2014

FOREWORD

ORGANIZING AND SCIENTIFIC COMMITTEES

DSIE'14 Organization and Scientific Committees welcome you to the Doctoral Symposium in Informatics Engineering of 2014. The organization of a scientific event, regardless of its small size, proved to be a difficult and complex task. We take from this experience great value and knowledge that will be with us for the future. It was with great honor that we accepted these functions in order to perform a scientific event successfully.

The success of this event is anticipated specially because of its contribution to the science community. With the effort of all students in the Doctoral Program in Informatics Engineering, it was possible to organize this event and therefore get the quality of work referenced above.

We would like to thank all the senior members of the Scientific Committee for their dedication and involvement in the DSIE'14. We would also like to thank the significant help of Sandra Reis, from the Informatics Engineering Department of FEUP, and to all the sponsors of the Doctoral Symposium of Informatics Engineering for their support and involvement to help DSIE'14 to be a reality.

And we also would like to thank all participants of DSIE'14.

Pedro Sousa and Rodolfo Matos (Organization Committee Chairs)
Benedito Junior and João Reis (Scientific Committee Chairs)
January 2014

CONFERENCE COMMITTEES

STEERING COMMITTEE

A. Augusto Sousa
Eugénio Oliveira

ORGANIZING COMMITTEE CO-CHAIRS

Pedro Sousa
Rodolfo Matos

ORGANIZING COMMITTEE

Álvaro Câmara
Benedito Bitencort
Danilo Jodas
João Reis
Leonel Dias
Mara Martins
Miguel Nabuco
Pedro Pinto
Pedro Sousa
Ricardo Gonçalves
Rodolfo Matos
Tiago Cunha
Vasco Miranda

SCIENTIFIC COMMITTEE CO-CHAIRS

Benedito Bitencort
João Reis

SCIENTIFIC COMMITTEE

Alexandre Valle de Carvalho
Álvaro Câmara
António Coelho
António Pereira
Benedito Bitencort
Carlos Pinto Soares
Danilo Jodas
Francisco Vasques
Henrique Lopes Cardoso
João Cardoso
João Pinto Ferreira
João Mendes Moreira
João Reis
José Magalhães Cruz
Leonel Dias
Mara Martins
Miguel Nabuco
Paulo Machado Portugal
Pedro Pinto
Pedro Sousa
Pedro Souto
Ricardo Gonçalves
Rodolfo Matos
Rosaldo Rossetti
Tiago Cunha
Vasco Miranda

SPONSORS

DSIE'14 – Doctoral Symposium in Informatics Engineering is sponsored by:

CONTENTS

INVITED SPEAKERS

SESSION 1 - MACHINE LEARNING & INTELLIGENT SYSTEMS

SESSION 2 - INFORMATION SYSTEMS & COMMUNICATION NETWORKS

SESSION 3 - SOFTWARE ENGINEERING & SECURITY SYSTEMS

SESSION 4 - SIMULATION SYSTEMS & HIGH-PERFORMANCE COMPUTING

INVITED SPEAKERS

INVITED SPEAKER

GONÇALO QUADROS

Gonçalo Quadros is the co-founder and CEO of Critical Software SA, an international software company, whose headquarters are in Coimbra, Portugal, with other offices in Lisbon, San Jose (US) and London (UK).

Gonçalo Quadros is 42 years old, has graduated in 1987 in Electrical Engineer (Computer Science branch) and has worked in industry and R&D prior to obtaining his PhD in Computer Science from the University of Coimbra, in 2002. His PhD work focused on Quality of Service over IP networks. In his academic life, Gonçalo Quadros taught at University of Aveiro and University of Coimbra (mathematics, computer and data communication subjects), authored more than 40 papers and was a reviewer for several conferences and magazines.

Gonçalo Quadros co-founded Critical Software in 1998. In that year, he won the best business plan award from the National Entrepreneurs Association (ANJE). He was the responsible for launching and consolidate two Company Business Units – Telecommunications and Networking and Enterprise Solutions –, while assuming responsibilities of project and technical management on some of the most important projects. He was also in charge of business development and team management. As of the beginning of 2005, Gonçalo Quadros is the Chief Executive Officer of Critical Software and maintains currently the responsibilities in the business development field.

INVITED SPEAKER

PAVEL BRAZDIL

Pavel Bernard Brazdil is originally from the Czech Republic, where he obtained a degree in Electrical Engineering, in 1968. In 1981 he finished his Ph.D. degree in Artificial Intelligence at the University of Edinburgh. Later, in 1982 he moved to Portugal to lecture at the University of Porto (Faculty of Economics). Since 1998, he holds the position of Full Professor at this Faculty.

He has lectured subjects at both under-graduate and post-graduate levels. He has supervised over the years 10 Ph.D. students, whom have completed their studies and still other two that remain in progress. Simultaneously, he has supervised 16 M.Sc. students.

Between 2007-11 he was the Director / Scientific Coordinator of R&D Unit LIAAD-INESC Porto L.A. LIAAD and its researchers are well known both in Europe and world-wide for their activities and it has a regular influx of visitors and post-graduate students from abroad. He has more than 20 years of experience in coordination, as he was one of the founder of another R&D Unit (LIACC) in 1988.

Currently, his research activity is focused mainly in sub-areas of Artificial Intelligence, namely machine learning, data mining, meta-learning, text mining, information extraction, adaptive modeling, among others. His Ph.D. work was carried in the area of Machine Learning and was innovative, since there was relatively little done in this area. The area underwent major expansion later over the last 30 years.

Pavel Brazdil contributed to the development of the area of Machine Learning also through his publications. He has published more than 100 articles. Furthermore, he has (co-)authored/edited 5 books by major publishing houses (Springer, Kluwer) and he is one of the authors of the book on Metalearning, published by Springer in 2009. He has also (co-)edited 2 special issues of Machine Learning Journal (MLJ) and published various articles in it. A considerable number of publications are referred to on ISI (28) and several of them have dozens of citations. He is a member of editorial boards of several international journals, including the prestigious Machine Learning Journal.

Pavel Brazdil contributed to the development of the area of Machine Learning also by his involvement in organizing workshops and conferences. Overall, he has organized more than 10 conferences or workshops in the areas related to Machine Learning and Data Mining. He has served also on program committees of many top conferences in the area. Pavel Brazdil has promoted the emergence of the sub-area of Meta-Learning and more recently, Planning to Learn. The first topic was covered in two European projects (StatLog, Metal). He is a member of Portuguese Association of Artificial Intelligence and has served as its President for two terms (2000-03). Since 2008 he has the honour to be one of the ECCAI Fellows.

SESSION 1

MACHINE LEARNING & INTELLIGENT SYSTEMS

Towards an Industrial Agent Oriented Approach for Conflict Resolution

João Reis

A Centralized Approach Using Multi-Agent Systems For Air Traffic Control

Álvaro Câmara and Daniel Silva

Support Vector Machine-based Method for Navigation of Mobile Robots through Paths in Plantations

Danilo Jodas

Sentiment Analysis on Twitter's Portuguese Language

Tiago Cunha

Towards an Industrial Agent Oriented Approach for Conflict Resolution

João Reis [1]

[1] Faculty of Engineering of the University of Porto. Rua Dr. Roberto Frias, s/n, 4200-465,
Porto, Portugal

jpcreis@fe.up.pt

Abstract. The agent paradigm is being strongly discussed for the past few years, mainly when addressed to practical and real world issues. The industry domain, and specifically the production system context, has revealed to be suitable for the use of Multi-Agent Systems, and along with it, some artificial intelligence concepts applied to process optimization [1]. The approach presented in this paper makes use of the shop-floor equipment representation by means of the agent concept, and mark-up languages for document encoding as XML to create a simulated environment. The purpose of all those aspects is studying the shop-floor dynamics, regarding the reliability and effectiveness in using this kind of paradigm in production systems. The present approach aims for the optimization in specific stages of the production system, like product ramp-up, scheduled maintenance and unscheduled maintenance. These different phases revealed to be very time consuming, and thus, task driven communication and negotiation strategies will actually open a door towards a whole new holistic perspective about shop-floor interactions. In this paper, a simple welding scenario is presented, looking forward to foster and leverage the use of, firstly, Multi-Agent System's concept within the industrial domain, and secondly, negotiation strategies to solve and handle conflict issues.

Keywords: Multi-agent System, Agent Negotiation, Agent Collaboration, Modelling, Production Systems, Manufacturing Execution System

1 Introduction

Situations like absence of equipment visibility at the shop-floor level, non-existing inter-equipment communication and lack of collaborative capabilities are just few challenges that nowadays European Industry have to deal with. Most of the times, these kind of problems lead not only to inefficiencies in the very early stages of the production system, which involve the product ramp-up – since the production is initiated, some equipment need to be calibrated until some quality parameter is not met - , but also lead to high costs when equipment requires maintenance. In this latter case, there are two different situations that can be observed. On one hand, a problem in an equipment certain equipment is identified but it can still operate at the shop-floor level, and therefore, maintenance can be schedule in further future. On the other hand, if a severe problem occurs, an equipment might require an immediate maintenance forcing all the production system to stop. All these different problems represent the nowadays challenges that need to be tackled in order to minimize production costs, maximize the product outcome, and consequently improve the competitiveness on the industry world.

All the background provided for this work is based on the I-RAMP3 European Project, in which all the shop-floor equipment is encapsulated by a technology that aims to extend its communication capabilities and add intelligent functionalities. This way, external software components will have the capability to communicate with those devices, being able to create a runtime representation of the shop-floor workflow.

Based on the previous description, the purpose of this paper is to expose an agent oriented approach applied to the industry domain, in which a simple Welding scenario is explored. This agent approach aims to create a representation of each equipment on the shop-floor level, like the I-RAMP3 Project does, but using instead an agent paradigm. Therefore, collaboration and negotiation strategies were explored in order to tackle shop-floor workflow optimization issues, like which equipment should be used for a specific required task and provide the best quality results.

The main goal of this approach is to explore and study the use of Intelligent Agent paradigm applied to the Industry domain. It aims to replicate the shop-floor equipment along with its dynamics and interactions and to explore the possibility of using this paradigm in a real industrial environment. To do so, the JADE platform was used to model all the shop-floor equipment, from the lower capacity devices like a temperature or humidity sensors, to the Manufacturing Execution System used to plan and coordinate the whole shop-floor. Taking advantage of the JADE functionalities, the Contract-Net protocol, along with specific JADE behaviours, were used to model the agents' interaction and the dynamics associated with the industry environment [2].

Section 2 is the first section of the paper, and it reviews some of the work that have been made in the field, which constitutes a very good basis for this approach exploitation. Section 3 presents in detail the proposed agent based approach, explaining the actors and its interactions. Section 4 shows some experimental results taking into account a simple Welding scenario, in which a set of sensors and a welding machine are used. In Section 5, a discussion is raised upon previous experimental results, concluding with few remarks and future work possibilities.

2 Related Work

This section will present two different frameworks that aim to apply the agent paradigm into the industry domain, in terms of reconfiguration and agility of the manufacturing systems.

The first one is named MetaMorph II, and is an agent based architecture that aims to integrate different manufacturing activities like design phase, planning scheduling, etc, and allow the system reconfiguration. This architecture is oriented to distributed intelligent design and manufacturing that takes into account entities like suppliers, customers and partners for extended-enterprise issues [3]. However, this architecture is more oriented to the system adaptability rather than system configuration and reconfiguration [4].

The second one is called AARIA (Autonomous Agents for Rock Island Arsenal), and is an agent architecture that is mostly concerned about the system's design. It is composed by collaboration models that are requirement driven, rather than using the collaboration capabilities for designing and specifying the system [5].

3 Agent Oriented Approach

3.1 Industrial Overview

As previously said, one of the purposes of this approach is to create an agent representation for each equipment on the shop-floor. To correctly understand the meaning of each agent, we need to previously be aware, in a simplified way, of the industrial panorama on two different stages of importance: 1) information system level that is oriented to the industrial domain and aims to set up the shop-floor configuration and plan the workflow according to the product specifications, called Manufacturing Execution System (MES); 2) The different types of equipment that can compose a shop-floor production system. The first one aims to manage and control all the shop-floor equipment, like receiving the product specifications and translate them into tasks that can be delegated among all the available equipment on the shop-floor level. The second one is directly related to the machines that are displaced throughout the shop-floor. In this latter, two different types of equipment were identified: high capacity devices and lower capacity devices. The high capacity devices are equipment with high capabilities of processing and memory that can control other high capacity devices or different sensors and actuators. The lower capacity devices are just sensors and actuators that need to be controlled by another entity, due to the lack of processing and memory.

3.2 Agents' Description

For a better understanding of this approach's purpose, the concept of agent needs to be explained and clarified. The definition of agent varies in ranges of context, assuming different functions and purposes in areas like philosophy, sociology, economy, law, among others. Despite those domains, the definition that seems to suite the industry context, lies in the following: *"A computer system that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objective"* [6]. An agent should have the a sense of autonomy, acting independently from human intervention, should also have the ability to interact with other agents and the environment, react according to the environment changes, percept from it, assume a pro-active posture in acting by its will, and should be a continuous running process.

Taking into account the Industrial Overview presented in the previous subsection, the was mainly influenced and inspired by the I-RAMP3 project, a set of agents was identified to fulfil the intended representation of the industrial environment: *Device Agent*, *Sensor & Actuator Agent*, and *MES Agent*. The *Device Agent* is a direct correspondence of high capacity devices and it aims to replicate the behavior of the equipment that have the high processing and memory capabilities, and is intended to control / manage other *Device Agents* or *Sensor & Actuator Agents*.

Sensor & Actuator Agent is the representation of the sensors and actuators that exist on the shop-floor. Since these kind of equipment are lower capacity devices, to operate on the production system they need to be controlled / managed by *Device Agents* or the *MES Agent*.

Finally, the *MES Agent* is intended to partially replicate the behavior of the *Manufacturing Execution System*. This partiality is based on the equipment shop-floor planning, in which tasks are delegated to *Device* and *Sensor & Actuator Agents* to execute according to its specification.

3.3 Agents' Interaction

One of the things that is inherent to a Multi-Agent System is the interaction among all the agents. This way, a very well-structured communication needs to be specified in order to guarantee a correct interpretation of what each equipment is capable to operate on the shop-floor level, all the tasks that need to be delegated and executed, and even additional functionalities that can be independent from the physical execution, like a service that provides a machine's information on a given period. Therefore, a set of eXtensible Markup Language (XML) documents was developed to fulfil all the communication needs and promote a more reliable interaction between agents. The XML format was chosen because it is both machine-readable and human-readable, is a very well accepted standard for document encoding [7], and there are several tools that can translate the information present on a class model into an XML-based file, and the other way around. Hence, the following set of documents were defined: *Self-Description Document* (*SDD*); *Task Description Document* (*TDD*); *Task Fulfilment Document* (*TFD*) [8].

Self-Description Document

The *Self-Description Document* contains the equipment basic information and all the tasks that can be executed on the shop-floor level. It aims to be sent to all the system entities (*MES Agent*, *Device Agent*, *Sensor & Actuator Agent*) when an agent enters the network of devices and is the basis for all the other documents to be exchanged during the agents' interaction.

Task Description Document

The *Task Description Document* can be seen as a task definition to be delegated among shop-floor equipment, and de executed according to the desired sender parameterization. This document makes use of the task structure provided by a certain device's SDD, and then specifies the parameterization to be executed by other agent, along with task duration and starting time.

MES Agent and *Device Agents* are the entities that can delegate tasks to others, and, on the contrary, *Device Agent* and *Sensor & Actuator Agent* are the ones how can receive task requests for further shop-floor execution.

Task Fulfilment Document.
The *Task Fulfilment Document* is the direct response to the *TDD*, and it uses the same structure of it. The receiver agent has the responsibility to update the sender desired parameterization of the TDD with the parameterization that is possible to be executed on the shop-floor level into the TFD, taking into account the receiver physical specification.

3.4 Agent Negotiation

When coordination and collaboration is intended to make part of the agents' dynamics in a Multi-Agent System, negotiation strategies need to be considered. The purpose of Negotiation in this specific approach is related with agent's cooperation to meet local and global goals, in which agents must act as a group and decide who executes what according to cost and utility functions.

In this case, it was used the Contract-Net Negotiation protocol to deal with resource allocation conflicts. Simply describing it, the Contract-Net protocol always need as least two different entities: the *ContractNetInitiator* and *Contract-NetResponder*. The first step of the protocol is made by the *Initiator*, sending a Call for Proposal (CFP) message to the *Responder*, in order for the latter to provide a Proposal according to what was specified previously on the proposal call. Finally the *Initiator* has the main role to analyze the proposal using a *Utility Function* deciding if it should accept or refuse the proposal. The idea of this approach is to use this protocol to deal with conflicts where two or more agents want to delegate a task to the same agent, and they need to agree upon who will make use of the task delegation.

For the evaluation of documents (both TDD and TFD) two different functions were built to quantify the cost associated to a task execution by a specific agent, and to measure the utility of a task execution by a different agent. Therefore, a *Cost Function* and a *Utility Function* were defined, along with a simple *Threshold Function* that calculates how much an agent is willing to negotiate a given task execution.

Cost Function.
Cost Function is intended to quantitatively evaluate a function in terms of execution impact for the production system. In equation (1) is presented the main parameters that should be taken into account to calculate a value that represents the effort applied to a task execution.

$$\frac{Ca}{Cr} + \left(\frac{St}{T}\right)^{Cr} \quad (1)$$

For the task execution cost to be calculated, a set of arguments need to be specified: *Cr*: required components of a task; C*a*: available components associated with a task; *St (serviceMinTime)*: value that describes the amount of time an equipment worth to execute a certain task on the shop-floor; *T:* task duration.

Utility Function.
Utility Function is basically a quantitative measurement of how useful a task execution can be, taking into account the differences between the requester's ideal task execution and its possible execution according to the specifications of the requested agent.

$$\frac{10}{[\sum_{i=0}^{A} diff\ i] * 0.1 + 10} \quad (1)$$

Taking into account Equation 2, *A* is the set of arguments present in all components that compose a certain task, and *diff i* is the difference between the same argument value defined on TDD and TFD. If the difference is 0, the utility will be 1, and in other cases, the utility will be any number between 0 and 1.

Threshold Function.
Equitation (3) is a simple equation which is intended to define how much an agent should be willing to negotiate for a task execution with others. In this particular case, parameter *P* assumes the value of 10, representing the maximum value it should be willing to negotiate.

$$P * Utility \quad (2)$$

4 Experimental Results

The main purpose of this section is to gather all the concepts and ideas described in the previous sections, and apply them to a very simple case for a better and easy reader's understanding of the system. It aims to provide not only a holistic overview of the approach, but also to frame all the concepts in a specific context, clarifying the main role of the proposed approach. Also this case study was inspired and based on the expertise of the all partners involved on the I-RAMP3 project, in which a Welding scenario takes an important role not only for project's requirements, but also as a basis for final project's demonstration purposes.

4.1 Case Study: Sensor Negotiation

Scenario Dynamics.

The first step to be taken towards sensor negotiation is to define the agents that will make part of scenario. Hence, we must defined two different entities that aim to operate on the shop-floor level - *Welding Machine* and *Metrology Station* - that implies the use of two instances of *Device Agents*, along with the use of one Mote that requires a representation of one *Sensor & Actuator Agent*, with the capability of providing three different types of measurement (task composed by three different components) at the same time – Temperature, Humidity and Camera.

For a simple and clear explanation of the scenario, **Fig. 1** will be used as a reference. It has the representation of two *Device Agents* (*Welding Machine* and *Metrology Station*) and one *Sensor & Actuator Agent* (Mote with three different sensors integrated). Before starting to make use of **Fig. 1**, an initial interaction between agents need to be explained. It was previously described that when an agent enters the network (environment where all the devices can virtually see each other), they needs to broadcast its internal information and all the tasks that they can execute. Therefore, taking into account the entities of the present scenario, when the two *Device Agents* and the *Sensor & Actuator Agent* step into the network, they need to generate a document that describe themselves (SDD), and then send it to all the devices that make part of the network. The purpose of sending this document to all the network components, is to make them aware of what the other ones can do, and therefore, giving the possibility to locally make decisions about which agents are suitable to collaborate with.

After this registration process on the device network, everything is aligned to start the task delegation process, where agents create documents (TDDs) where they can specify the desired conditions of a certain task execution. Thus, it can be seen from **Fig. 1** that the initial message sent from the *Device Agents* to the *Sensor & Actuator Agent* is a *Task Description Document,* requiring the execution of a specific. This step corresponds to the Call for Proposal message on the Contract-Net Protocol, where agents send the desired parameterization for task execution. Consequently, and regarding this scenario, the Mote should be able analyse the two received TDDs, and do two different things in each case: Generate a TFD based on TDD; Calculate the cost of TFDs.

The first thing is basically to compare what was defined in the TDD with what the Mote can actually execute on the shop-floor (SDD specification), and generate the TFD accordingly. The second thing is it to make use of the defined *Cost Function* to estimate the price that a *Device Agent* needs to pay for the task defined on the TFD. This value is a measure used to compare the effort that needs to be applied for different tasks by different devices, and to see if it worth the use of that equipment or it makes sense to reject that specific task execution. After those two steps are completed, the Mote device needs to send the TFD back to the *Welding Machine* and to the *Metrology Station*, along with the cost of each TFD execution.

For each *Device Agent* to measure how suitable a task response (TFD) is comparing with the desire task execution (TDD), it needs to make use of the *Utility Function*. This function, as previously explained, measures the distance between the ideal equipment operation and what is can really execute on the shop-floor level. It aims to quantify if the proposal sent from the *Sensor & Actuator Agents* is useful or not for the task execution to be used. Therefore, each *Device Agent* should compare the task execution usefulness with the calculated *Threshold* value (which is based on the task execution utility), and analyse if it worth to pay for its execution, and accept or not the *Sensor & Actuator Agent*'s proposal. In the case of both *Device Agents* accept the proposal, a conflict is identified, and the *Sensor & Actuator Agent* is responsible to increase the costs of the task execution to see how's willing to pay the higher value.

This market-based negotiation strategy aims to take the maximum advantage of the task execution requesters, since in cases where the demand is higher than supply, the product prices will increase and maximize its profits. In this case, since we are talking about collaboration, these prices are not actually to be paid, but just a measure to determine whose component will make use of the task execution.

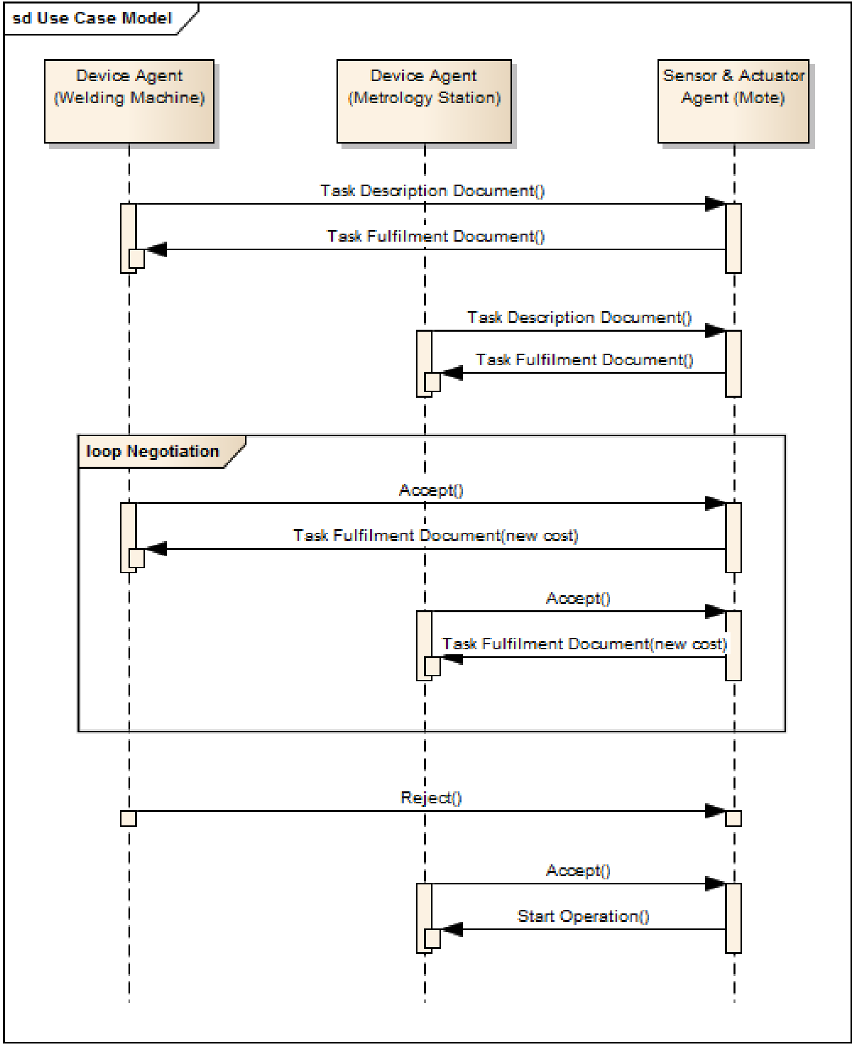

sd Use Case Model
Device Agent (Welding Machine)
Device Agent (Metrology Station)
Sensor & Actuator Agent (Mote)
Task Description Document()
Task Fulfilment Document()
Task Description Document()
Task Fulfilment Document()
loop Negotiation
Accept()
Task Fulfilment Document(new cost)
Accept()
Task Fulfilment Document(new cost)
Reject()
Accept()
Start Operation()

Fig. 1. Simple Negotiation Sequence Diagram

Finally, when the task execution cost is higher enough for only one *Device Agent* paying for the execution, a message will be sent informing about the success of task execution allocation, and it will start to operate in the specified time, during the specified task duration.

Simulated Scenario.

For full and complete understanding of the information exchanged in the negotiation process, a simple example will be presented using the previous explained equipment, protocols and information documents. As a basis for this exposure, **Table 1** shows the information that was specified regarding the task execution.

It can be seen that the Task Type is *Measurement*, and it can use three different types of *Components* at the same time: Temperature, Humidity and Camera. From the table, it can also be seen the information on the task level, *ServiceMinTime* and *NumberComponents*m and on the sensors level, a set arguments used to parameterize the operation conditions of the Mote.

In the initial negotiation phase, *Welding Machine* and *Metrology Station* should create a TDD and send it to the Mote device. In this paper, we will not focus on TDD and TFD definition, but on the process of reaching an agreement of whom will make use of the task execution. Therefore, only the differences between TDD and TFD will be considered, rather than the whole task description. *Welding Machine* wants to make use of all the three different sensors, and the *Metrology Station* wants to make use of only one sensor: Camera. From this point, we can assume that *Welding Machine*'s TDD should have the information of the three sensor parameterization, and the *Metrology Station* should have only one sensor parameterization.

With the present information, we can start specifying some arguments for the *Cost Function* to be used by the *Sensor & Actuator Agent* for each TDD received. Considering the *Welding Machine*, and taking into account equation (1), we can define Ca = 3, Cr = 3, St = 100 and T = 500. Considering the *Metrology Station*, we can define Ca = 3, Cr = 1, St = 100 and T = 20. As can be seen, there are two differences on the function's arguments. One of them is the number of components required - *Metrology Station* only requires one sensor - and task duration - *Welding Machine*'s task is required to be longer than the *Metrology Station* one. Hence, the cost associated to the required task from *Welding Machine* is 1.01 and from *Metrology Station* is 8.0. The big difference between costs is due to two major aspects: Number of sensors consumed and Task duration.

As explained in the previous sections, considering the Mote, it is more costly to use only one sensor rather than all of them, because we are avoiding others to use the full capabilities of the device, and it is also infeasible to execute a task with duration lower than the *ServiceMinTime*, for energy consumption reasons. Regarding the two TFDs that should be generated by the *Sensor & Actuator Agents*, we will assume there are two differences comparing with the *Welding Machine* TDD (two different task arguments cannot be fulfilled by the mote specifications), and no difference comparing with the *Metrology Station* (the task can be executed without restrictions).

Therefore, *Sensor & Actuator Agent* should send the TFDs to the corresponding entities, and wait for either an acceptance or rejection response. On the *Device Agent*'s side, an utility for each received TFD should be calculated. To do so, we only need to know the number of sensor arguments that can be parameterized, and the sensors required to be used, since the *Utility Function* makes use the differences between what is ideal to be executed, and what can actually be operated on the shop-floor. Thereby, TFD utility for the *Welding Machine* is 0.909, and for the *Metrology Station* is 1. The calculated utility for the first equipment is only 0.909 because, as previously depicted, there was not a full matching between the presented TDD and the proposed TFD.

Table 1. Sensor & Actuator Agent - Measurement Task Description

Type	**Service Min Time**	**Number Components**			
Measurement	100	3			
Components	**ID**	**Description**	**Arguments**		**Unit**
	Temperature	Temperature Sensor	Min	-55	ºC
			Max	200	ºC
			MinError	10	%
			MinResponseTime	55	Ms
	ID	**Description**	**Arguments**		**Unit**
	Humidity	Humidity Sensor	Min	0	%
			Max	100	%
			MinError	20	%
			MinResponseTime	55	Ms
	ID	**Description**	**Arguments**		**Unit**
	Camera	Camera Sensor	maxFrameRate	24	Img/sec
			minLatency	100	Ms
			MinError	20	%
			MinResponseTime	55	Ms

The next phase of the negotiation protocol is the first step of the negotiation loop presented in **Fig. 1**, in which *Sensor & Actuator Agent* iteratively increases the cost of the task execution to determine which *Device Agent* is willing to pay the higher price for the task execution. As explained before, the cost value is just representative and with no influence. For instance, if a *Device Agent* accept a task execution with a cost value of 8.5, anything will change or influence in further operation, since we are dealing with a collaborative environment, and not a competitive one. Therefore, the *Threshold* value that *Welding Machine* and *Metrology Station* are willing to pay for a task execution is 9.09 and 10, correspondingly. Whenever the *Sensor & Actuator Agent* receives the acceptance messages from the two *Device Agents*, no agreement is reached, and it will increase the cost of each function in 1 unit. Hence, in the next iteration, task execution costs will assume the values of 2.119 and 10.25, and taking into account the previous thresholds, the *Device Agent* representing the *Metrology Station* is not willing to pay for the task execution, and will send a *reject* message to the *Sensor & Actuator Agent* saying that it is no longer interested in the task execution. On the other hand, since the new *Welding Machine* task execution cost does not exceeds the calculated threshold, it will send an acceptance message to the Mote.

Finally, and after *Sensor & Actuator Agent* checks that only one *Device Agent* is willing to pay the cost associated with the task execution, it will send a *Start Task Operation* message informing about the success of the negotiation.

5 Discussion

Throughout the whole paper, new concepts and ideas were explored with the intention of improving some of the nowadays difficulties in the industry domain. Regardless of the simple example presented in the section 4, the strategies applied to this context, along with the agent paradigm and well structure communication processes, proved to be a good and reliable approach. Nevertheless, more industrial scenarios and validation processes are required and need to be exploited, until this approach reach the necessary consistency to gain credit among the industrial reality.

One of the most important advantages of the presented concepts, is undoubtedly the decentralized approach, which seems to verify the fault tolerant property. This means that in case of sudden equipment fail, the network of equipment will maintain its communication and collaboration activities, avoiding stopping the production system due to component dependency issues.

Another concept introduced in this paper, is the task driven communication, in which not only equipment execution on shop-floor level but also internal programmed services, are specified in XML-based format, and used to delegate responsibilities of providing certain results and operation according to precise specifications. This concept allow the automatic reconfiguration of equipment for shop-floor operation, in which, in some of industrial contexts, is made manually and reveal to be very costly and ineffective.

Obviously that when we talk about a complex and ambitious approach like this, the other side of the coin needs also to be revealed, and therefore discussed and explored. When we talk about industrial systems, latency is always an issue that needs to be tackled when exploring and developing new concepts to be applied. In this approach, as previously described, to deal with resource conflict a negotiation protocol was used. In an environment where there are dozens of devices communicating with each other, the Contract-Net protocol is used for negotiation based on market, and it can represent an undesirable network overhead.

The last issue that will be dissected in this section, is the additional computational resources that this approach involves. As said in previous sections, nowadays, there are components like sensors and actuators that don’t have a high processing and memory capability, and therefore, cannot be added more complex functionalities and communication means. Hence, investment on hardware components and physical architecture reconfiguration for the equipment’ complement of communication capabilities and intelligent functionalities, is a real restriction.

The next steps towards this specific industrial agent oriented are related with the use of learning techniques for negotiation improvement and development of predictive maintenance models. The first one lies in the threshold learning, in which the amount effort a device is willing to apply in negotiation can be adapted during the whole production execution, being the second one a more complex and challenging step, due to the fact that all the system's data must be correctly analysed and predictive models need to be created accordingly for further adaptation as runtime information is becoming available. This would support the maintenance personal with, firstly, know which part of a component needs to be maintained, and secondly, when a sudden fail is probable to occur.

6 Acknowledgements

This work was carried out with the support from the European Commission under the Seventh Framework Programmer, in the Integrated Project I-RAMP3 (FoF.NMP.2012-3).

7 References

1. Shen, W., Norrie, D. H., and Barthès, J.-P. A.: Multi-Agent Systems for Concurrent Intelligent Design and Manufacturing. Taylor & Francies, London and New (2001)
2. Bellifemine, F., Poggi, A., Rimassa, G.: Developing Multi-agent Systems with JADE. In : Intelligent Agents VII. Springer-Verlag Berlin Heidelberg (2001)
3. Shen, W., Maturana, F., Norrie, D. H.: MetaMorph II: an agent-based architecture for distributed intelligent design and manufacturing. In : Journal of Intelligent Manufacturing. (2000)
4. Ferreira, P.: An agent-based self-Configuration Methodology for Modular Assembly Systems. PhD Thesis, University of Nottingham, Nottingham (2001)
5. Parunak, H. V. D., Baker, A. D., Clark, S. J.: The AARIA Agent Architecture: From Manufacturing Requirements to Agent-Based System Design Integrated. In : Computer-Aided Engineering. (2001)
6. Wooldridge, M.: An introduction to multiagent systems. 2nd edn. John Wiley & Sons, Chichester, England (2002)
7. Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergeau, F.: XML 1.0 Specification., W3C (2008)
8. Peschl, M., Link, N., Hoffmeister, M., Gonçalves, G., Almeida, F. L. F.: Designing and implementation of an intelligent manufacturing system. In : Journal of Indutrial Engineering and Management. (2011)

A Centralized Approach Using Multi-Agent Systems For Air Traffic Control

Álvaro Câmara[1], Daniel Silva[1]
[1] Faculty of Engineering of the University of Porto,
Street Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
{up201300764,dcs}@fe.up.pt

Abstract. Air traffic has been increasing in the last decades, representing a major mode of transportation. The FAA (Federal Aviation Administration) and the aeronautical industries are both predicting a growth between 150% to 250% over the next two decades. In this paper, a multi-agent able to autonomously manage the traffic control system on an airport will be presented, through an agent that will assume the role of an air traffic controller, in a control tower of an airport. During the simulations we were able to avoid collisions between aircrafts in an airspace with densities up to 6.125 km ^{2} per aircraft.

1 Introduction

This study is motivated by the increasing growth of world air traffic, so in order to better understand the changes that global air traffic has suffered, only in the European continent by the year 2030 we will have an average growth of 2.8% per year. This means that on a typical day we will have 500-1500 new flights to be monitored and all this demand should be accompanied by new technologies, equipment and processes to assist the ATM (Air Traffic Management), otherwise we might face serious problems in the future [1].

With the constant evolution of aviation and its popularization as a means of transportation faster than others, for large distances, it became increasingly necessary to provide to all who use it greater safety and better efficiency in their processes. With the constant search for new technologies and equipment that can automate the processes of air traffic management (ATM), this project aims to provide new systems that can assist the ATM through the use of multi-agent systems.

In this paper we propose and demonstrate that the use of multi-agent systems for air traffic control can be effectively applied to solve the problems of detection and solving air conflicts, thereby increasing the safety of the air space and reducing the workload on the air traffic controllers.

Section 2 is an introduction to some of the entities that are present in the dynamics of air traffic control and the presentation of some of the proposals that use multi-agent systems to ATM. Section 3 describes the way in which this proposal was implemented and the used resources. In Section 4 we present the testing scenario and discuss the results. Finally Section 5 draws some conclusions and future work is presented.

2 Background

This section will discuss some needed concepts for a better comprehension of the dynamics of the Air Traffic Control (ATC) and some studies and proposals on the inclusion of multi-agent systems in ATC.

2.1 Air Traffic Control

The Air Traffic Control (ATC) is a service provided for maintenance and air traffic management through a series of procedures almost always associated with a set of technological devices such as radars, communications systems and weather systems. To provide this service to all who are using the aerospace (aircraft, balloons, and helicopters among others) the best possible information is necessary in order to have a safe flight [2].

The objectives of ATC are to avoid collisions between aircraft both in the flight phase as well as when they are maneuvering on the ground, maintaining an orderly and continuous flow of air traffic, providing advice and information useful for the safe and efficient conduct of flights and alerting appropriate agencies whenever an airplane has need of search and rescue services [3].

2.2 Air Traffic Controller

The air traffic controller (ATCO) is responsible for coordinating aircrafts both in the air space as well as in airports, in order to maintain continuous, orderly and safe airflow [4], through instruction and guidance to pilots to avoid collisions between aircrafts or objects on the maneuvering area.

Several studies indicate a growth in air traffic worldwide, as well as continuous increase in the workload for professionals in the air traffic control [5].

The increased workload is directly associated with the following factors: status of the airspace, equipment condition (design, reliability and accuracy) and the current state of the driver (age, experience). These parameters have a direct impact on the capacity of control of an airspace [6].

An excessive increase in the workload of these professionals tends to increase the amount of information exchanged between ATCO and the pilots and can thus cause a physical and mental exhaustion, contributing to a possible human error [7].

2.3 Conflict Detection and Resolution Systems

In an attempt to automate and increase the efficiency of some processes of the ATC, the scientific community has proposed numerous ways of inserting multi-agent systems in civil and military aviation. Among the various proposals, we highlight the centralized and decentralized approaches, taxiways management, management of air traffic flow manager and aid in a decision making process using multi-agent systems.

The authors [8] propose a centralized approach in which the air traffic management (ATM), can be solved using a simple iterative method called "Space-Time Flow" (STF). This solution can be deployed in a center of air traffic control, where through an agent you can look for possible collisions and propose multiple secondary routes, all free of collisions. Both the proposed solution by these authors as well as the solution presented in this article fall within the same range of solutions, but with some different approaches that can be better understood in the following sections.

The authors [9], believe that a decentralized approach could increase the efficiency of airspace, because in case of a possible mid-air collision the aircraft themselves could negotiate measures to prevent the accident. This article proposes the use of the Monotonic Concession Protocol (MCP) which has been shown to capture the ideas of negotiation involving relatively simple calculations and most importantly, this protocol behaves robustly, and doesn't allow being controlled by other agents in order to win some kind of advantage in the system. Other solutions involving this type of approach can also be found in [10] and [11].

The separation strategies are not only to ensure the safety of aircraft in the flight phase, but there is also the need to provide greater security to all the aircrafts underground maneuver and also to other vehicles that may transit on the taxiway.

Because of the similar characteristics between the management of taxiways with other scenarios such as urban traffic control and traffic control trains, we could replicate their solutions to the problems of managing taxiways. Where in both of them we find traffic restrictions in certain areas, intersections, holding points (light) and others. It is proposed by the authors [12] the use of multi-agent systems to achieve control of train traffic, specifically the resolution of conflicts in points of intersection of the railroads. In this approach agents are inserted in the crossings of railroads that manage the passage of trains (also represented by agents) to ensure the safety of all vehicles.

Unlike the approaches that deal with separation strategies the authors [13] propose the inclusion of multi-agent systems in air traffic flow management (ATFM). This proposal aims to identify in advance the congestion in the queues of aircraft held in the air, waiting for permission to land, preferring a deliberate retention of the aircraft at your airport of origin until it is possible to receive it "immediately" after its arrival at the destination. Unlike the current model that ends up holding the aircraft in the airspace above the airport of destination, thus enhancing the problems of safety of the aircraft, as the increased density of these airspaces, imply a greater effort of the air traffic controllers to ensure the safety of all aircraft.

The authors [14] are an example of a class of solutions that use multi-agent systems to propose solutions to the most diverse professionals who work in the ATM. The authors of this project have been designed to assist in the controller processes requests for landing, detect and avoid collisions, unlike [13], who use agents to identify in advance the congestion of the queues at airports and therefore try to reduce them.

3 Implementation

This section will look at the internal architecture of the ATC module developed in this project, the simulation platform used for the tests, the solution presented in this article and how it utility functions are used in the decision-making process of the ATCO agent.

3.1 Simulation Platform

Flight Simulator X (FSX) has been evolving since its first release thirty years ago, now in its tenth version, it became something greater than just a flight simulator, with a rich scenario and detail associated with a range of independent vehicles (land, air and sea) and a system that allows the creation of missions where the interaction between all components of the game [15] is possible.

The main reason for choosing this platform, was the fact that external applications can communicate with FSX, via an Application Programming Interface (API) called SimConnect. Another reason for choosing this simulation platform occurred because it is possible to find some desirable features in a simulator with a a well-developed physics system, taking into account factors such as aircraft aerodynamics, weight, amount of fuel and passenger, instrumental breakdowns, hydraulic and other crashes, as well as having open source code, making it possible to view and change it if necessary, and enjoyable graphics due to the use of DirectX10.

3.2 The Proposed Solution

In this approach, the ATCO agent centralizes all necessary activities for the resolution of air conflicts, leaving only the aircrafts on its domain to execute his orders. This technique has the main advantage of having a better view of the airspace as well as identifying problematic areas. However, its disadvantage lies in the fact that a failure in the central control device would paralyze the entire airspace over its domain.

To be able to monitor the aircrafts, the controller agent has a “Radar” at its disposal, in which you can see at all times the distance between the agents Aircraft. In this way two spheres of radius "R" and "2R" are simulated around the aircraft (see figure 1), being the "Red Zone" responsible for initiating the process of generating alternate routes if overtaken by other aircrafts, and the "Blue Zone" in the detection of new collisions generated by evasion maneuvers.

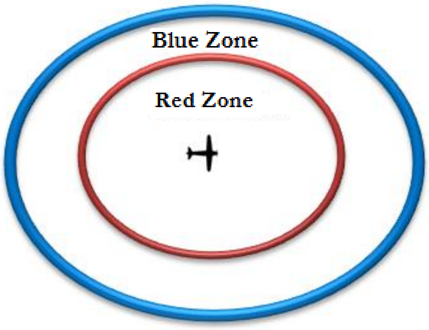


Fig. 1. Monitoring the areas around aircraft

By using these two areas it is expected that even during the evasive maneuvers, other possible collisions are not generated, since the controller agent uses a sixty seconds "time window", taking into account speed and direction, and can thus predict the generation of new collisions and avoid regions of higher traffic density.

Figure 2 presents an overview of the processes adopted in this proposal for the detection and resolution of aerial conflicts.

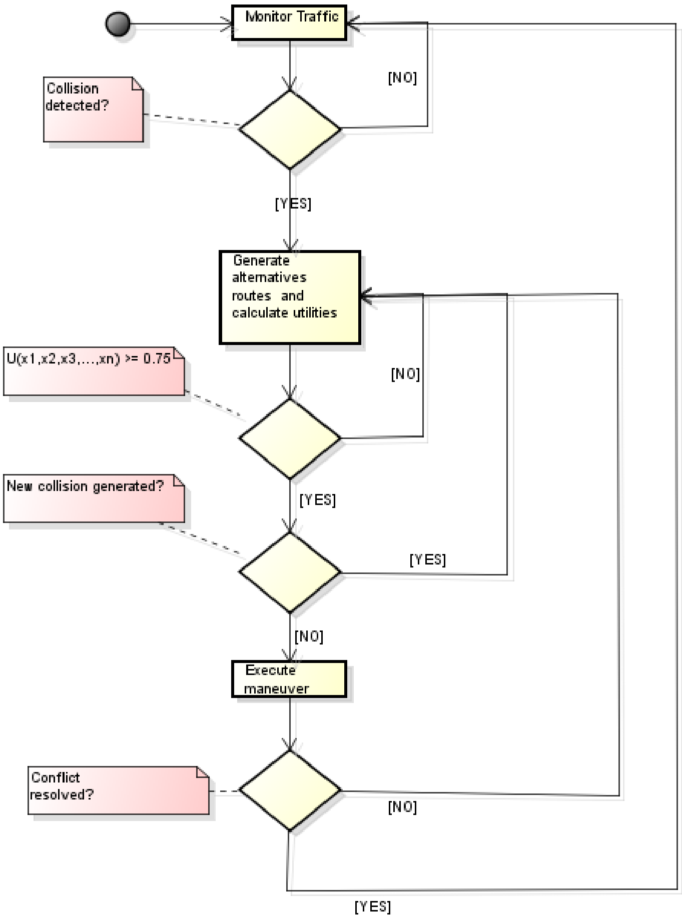


Fig. 2. Activity Diagram for the centralized approach

The process begins with the constant monitoring of the airspace by ATCO agent through the submodule Radar, were it constantly seeks some situations where one or more aircrafts end up invading their red zones.

After identified all possible collisions between two or more aircraft, it starts the process of generating alternative routes for all involved. The number of iterations in which these routes are generated must be limited in order to avoid an excessive number of proposals, due to the tight response time in which aircrafts must change their routes.

During the process of generating new routes, for each idealized new route, it is checked if the average of the utilities is greater than or equal to 0.75 and when performing this maneuver, one or more aircrafts will create a new condition of collision with the aircrafts present in the blue zone. After the fulfillment of these two conditions the aircrafts in conflict can perform their maneuvers, however if the maximum number of iterations in the development of new routes is exceeded, the set of routes that had the highest utility will be executed, whether or not they will generate future collisions within the blue zone.

After the controller agent sends the new routes containing a way point list that consists of a set of attributes (latitude, longitude, altitude and speed) he turns back to his observation status in the airspace until further possible collisions may reoccur.

3.3 Architecture of the ATC Module

In order to carry out air traffic control in [16], it was proposed the establishment of submodule "detection and conflict resolution"(DCR), considering the lack of effective management during the flight of the aircraft. To implement this solution (proposed in Section 3.2) it was necessary to change the submodule "Radar" in order to insert monitoring areas around the aircraft (red and blue zone), so that the new sub-modules could follow in real time all aircrafts when necessary and carry out the process of generating secondary routes.

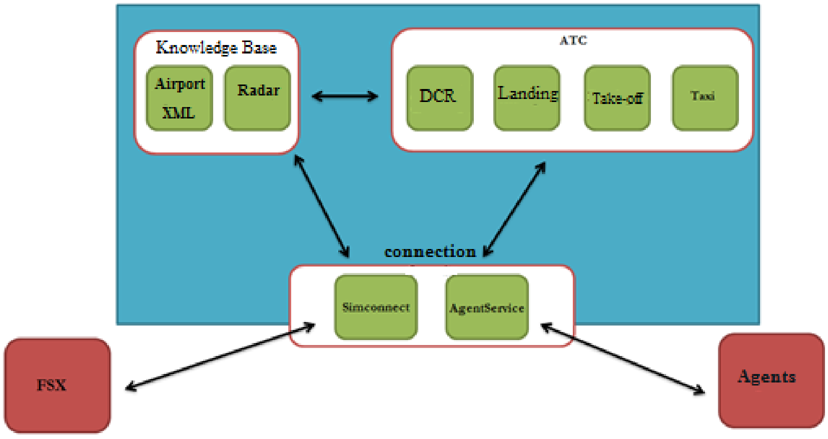


Fig. 3. Architecture of the ATC Module

The ATC module is divided into three parts, the same interconnected in order to perform the data exchange between the sub-modules. Below the description of each module will be held:

Knowledge Base: This sub-module contains all the information captured from the "real world", being used mainly by the ATC submodule to conduct their activities in managing the aircraft.

This sub-module contains 2 sections: "XML Airport" and "Radar", the first section being responsible for loading the information from the configuration file of the airport, save the data of the taxiway and runway and keep the table of best routes between taxiways. This information is loaded only at the beginning of the simulation and maintained statically during testing. The second section called "Radar" is responsible for monitoring the simulation time in all aircrafts, and store their navigation data (speed, altitude, direction, etc.). Their values change dynamically during the simulation, through Simconect which performs communication with FSX.

ATC: This sub-module is responsible for managing all activities related to the agents during the simulation. Within this sub-module are contained four sections: "Detection and resolution of conflicts", "landing", "Take-off" and "Taxi." Being the first section responsible for monitoring aircrafts in the flight phase in order to avoid possible air collisions. This section is primarily used by the centralized approach. The second section, is responsible for controlling the landing of the aircrafts, thus providing the release of the runways. Also, puts the aircrafts on hold and performs the restructuring procedures for the queue landing in case of a priority request.

The third section is responsible for freeing the use of the runway for takeoff and authorize their crossing, even if a second aircraft is in the landing process. Finally, the fourth section is responsible for managing the aircraft during taxi process.

Connection: In this sub-module communications between external entities to ATC module are performed. Within this sub-module 2 sections are contained: "SimConnect" and "AgentService". Being the first responsible for data exchange with FSX and the second for communication with agents.

3.4 Utility Functions

The Multicriteria Utility Theory (MAUT) is derived from the theory of utility associated with the decision-making process. Being used to define the importance of attributes with higher priority over lower priority by building mathematical function [17].

In this project the MAUT was used in the processes of decision making in order to verify that the new route generated by the avoidance maneuver fits the needs of the ATCO agent. These needs that can not always be met.

During the process of decision making, some measures were taken in order to avoid the continuous search for the best possible solution. To try to mitigate this situation, it was decided that after reaching at least the value of 0.75 in the additive utility function, the process of searching for better avoidance maneuver is interrupted and appropriate maneuvers are performed, another measure was to limit the number maximum routes generated

To calculate the additive utility function, the worst and best values for each attribute were initially established (see Table 1).

Table 1. Attributes and their expected conditions

Attributes	Worst	Best
Angle Bend (AB)	>= 25 degrees	0 degrees
Speed Variation (SV)	>= 30% of actual speed	Speed doesn't change
Traffic density of the new route (TD)	Half the testing aircrafts	0

Based on the parameter values given above, we can then calculate the intermediate utility values for each attribute, after the generation of a possible avoidance maneuver using equation 1.

$$Ui(x) = \frac{x\text{−worstvalue}}{\text{bestvalue} - \text{worstvalue}} \qquad (1)$$

For the next step you need to set the weight "K" of each attribute, this value depends mainly on the level of importance we associate to each attribute. We have situations where the bend angle is the most relevant criteria and others where speed variation is. For testing purposes we will use the weight values "K" for the ratio 6/1/3 AB, SV and TD respectively attributes.

With the results of equation 1 and their weights associated with each attribute, we can calculate the additive utility function (see equation 2). As the results contained in the range between 0 and 1, where zero is the expected worst and the one the best result.

$$\text{U}(\text{x1},\text{x2},\text{x3},\ldots,\text{xn}) = \text{k1U1}(\text{X1}) + \text{k2U2}(\text{X2}) + \cdots + \text{KnUn}(\text{Xn}) = \sum_{i=0}^{n} \text{K}\, iUi(Xi) \qquad (2)$$

4 Experiments and Results

4.1 Scenario

During the tests, the used airport was the Whidbey Island Naval Air Station (KNUW) and the used aircrafts were the Beech Baron 58 and the Cessna 172 skyhawk, both small and traditional in world aviation.

The table below describes the speeds in the taxiway and flight phase of the aircrafts during the simulation

Table 2. Speed of the aircrafts

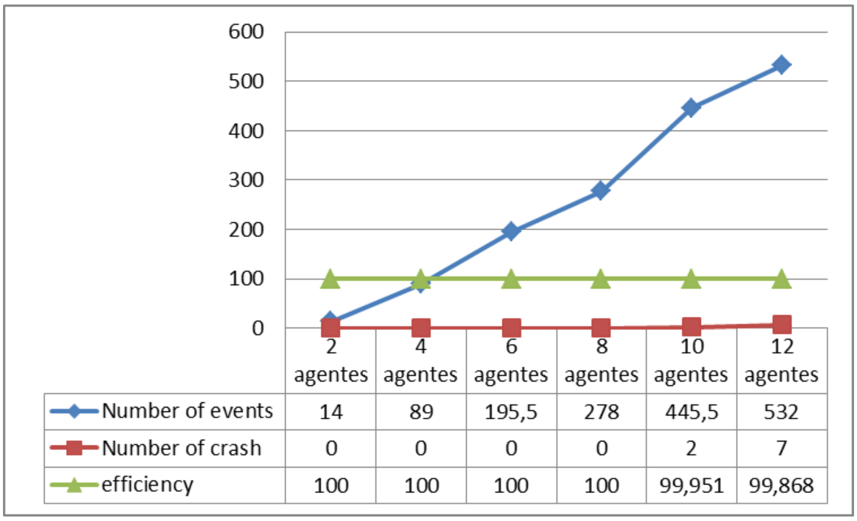
Aircraft	Taxiway Phase (knots)	Flight Phase (knots)
beech baron 58	10	151
cessna skyhawk 172	10	128

During the simulations in KNUW with a varied number of aircraft (2, 4, 6, 8, 10 and 12.). Being the same proportion of cessna skyhawk and beech baron 58. When starting the simulation aircrafts will make a request to the ATCO agent for takeoff and will enter the taxiway phase. After the first request the countdown begins.

After the aircrafts reach the flight phase, they are routed to a nearby airport area, with approximately 49 km^2 (3945 meters sphere radius), with the goal of generating the highest traffic density possible in a small region, thus leading towards greater occurrences of air conflicts. If an aircraft leaves the test region, it is randomly generated a new route within the respective zone.

Within the test area if two aircraft are less than 3600 meters apart (occurrence of an event), the process of conflict resolution starts. During the simulation, it will only be counted as a collision event if the distance between the aircraft is less than 20 meters. However, the aircrafts involved in these events will not be excluded from the simulation, thus keeping a fixed number of aircrafts throughout the tests that will run for 10 hours straight.

Figure 4 shows the mean number of possible collisions generated in a period of 1 hour, the number of collisions occurring during the 10 hours of testing and the efficiency of the centralized approach for different loads of traffic flow.



	2 agentes	4 agentes	6 agentes	8 agentes	10 agentes	12 agentes
Number of events	14	89	195,5	278	445,5	532
Number of crash	0	0	0	0	2	7
efficiency	100	100	100	100	99,951	99,868

Fig. 4. Simulations Results

Throughout the tests it was possible to guarantee an efficiency of 100% up to a maximum density of 6.125 km ² per aircraft. With the use of higher densities than this, it was realized that the workload of the ATCO agent almost doubled comparing to the number of events between tests 8 to 12 agents, which directly contributed to the first loss of efficiency of the centralized approach .

5 Conclusion

With the estimated increase in air traffic around 150% to 250% over the next two decades there was concern by the government agencies regarding the workload of air traffic controllers, considering that this profession is usually already quite exhausting.

We proposed to implement a simple conflict resolution system using multi-agent systems, where the role of the air traffic controller was accomplished by an autonomous agent. During the testing it was possible to meet a minimum safety requirement up to a maximum density of 6.125 km ² per aircraft, demonstrating that the use of multi-agent systems can be a potential technique to be applied in future proposals for solutions to air traffic control.

As future work of this project, a decentralized approach where the aircrafts themselves in conflict come into a negotiation process to resolve their conflicts in routes may be applied. Another possibility is a hybrid approach between the centralized and decentralized approaches, where during the negotiation process of the decentralized approach agents can consult the ATCO agent to aid in the decision-making process.

Acknowledgment

The first author is supported by CAPES (Coordination of Improvement of Higher Education Personnel) with reference doctoral fellowship BEX 0646/13-0.

References

1. EUROCONTROL, “EUROCONTROL,” 10 01 2013. [Online]. Available: http://www.eurocontrol.int/sites/default/files/content/documents/official-documents/forecasts/long-term-forecast-2010-2030.pdf. [Access 12 01 2013].
2. E. L. Ray, 12 10 2012. [Online]. Available: http://www.faa.gov/documentLibrary/media/Order/ATC.pdf. [Access 12 10 2012].
3. ASA, “ASA,” 17 02 1984. [Online]. Available: http://www.asa.cv/index.php?option=com_content&view=article&id=155&Itemid=203. [Access 13 10 2012].
4. G. M. Camargo, Controle de Pressão Seletiva em Algoritmo Genético Aplicado a Otimização de Demanda em Infra-Estrutura Aeronáutica. Tese de Mestrado, 2006.
5. A. Majumdar e W. Y. Ochieng, “The factors affecting airspace capacity in Europe: A Framework Methodology based on Cross Sectional Time-Series Analysis usingSimulated Controller Workload Data,” em *EUROCONTROL DAS/AFN Division*, Brussels, Belgium, 2002.
6. A. Majumdar e J. W. Polak, “A FRAMEWORK FOR THE ESTIMATION OF EUROPEAN AIRSPACE CAPACITY USING A MODEL OF CONTROLLER WORKLOAD,” em *Selected Proceedings of the 9th World Conference on Transport*

Research, Seoul, Korea, 2001.

7. M. L. Pogianelo e C. Müller, "ESPAÇO AÉREO BRASILEIRO - ANÁLISE DA CARGA DE TRABALHO DOS CONTROLADORES DE TRÁFEGO AÉREO POR MEIO DE SIMULAÇÃO "FAST TIME" - ESTUDO DE CASO NA TERMINAL RECIFE.," em *VIII SITRAER - simpósio de transporte aéreo*, Rio de Janeiro, 2008.

8. Y.-J. Chiang, J. T. Klosowski e C. Lee, "Geometric Algorithms for Conflict Detection/Resolution in Air Traffic Management," em *Conference on Decision and Control - CDC*, Atlanta, 1997.

9. S. Wollkind, J. Valasek e T. R. Ioerger, "Automated Conflict Resolution for Air Traffic Management Using Cooperative Multiagent Negotiation," em *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Rhode Island, 2004.

10. J. C. Hill, J. F. R. J. K. A. R. . L. F. e W. C. S. , "A cooperative multi-agent approach to free flight," em *roceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, New York, 2005.

11. K. J. P. M. K. D. B. C. L. e J. S. B. M. , "System performance characteristics of centralized and decentralized air traffic separation strategies," em *Fourth USA/Europe Air Traffic Management Research and Development Seminar*, 2001.

12. J. A. Guerrero, R. R. Gudwin, L. A. Domínguez e F. A. Gomide, "CONTROLE DE TRÁFEGO DE TRENS UTILIZANDO SISTEMAS MULTIAGENTES," em *Agent's Day – CBComp*, Itajai, Santa Catarina, Brasil, 2002.

13. M. V. Dib, L. Weigang e A. C. Melo, "Abordagem de Balanceamento de Negociação," *IEEE LATIN AMERICA TRANSACTIONS VOL. 5, NO. 5,* 09 2007.

14. H. Hexmoor e T. Heng, "Air Traffic Control and Alert Agent," em *Proceedings of the fourth international conference on Autonomous agents*, New York, NY, USA, 2000.

15. Microsoft, 20 12 2012. [Online]. Available: http://www.microsoft.com/brasil/games/fsx/default.htm. [Acesso em 20 12 2012].

16. D. C. Silva, "Cooperative Multi-Robot Missions: Development of a Platform and a Specification Language. Tese de Doutorado," 2011.

17. J. E. Teixeira e M. C. Belderrain, "APLICABILIDADE DA TEORIA DE UTILIDADE MULTICRITÉRIO (MAUT) NA SELEÇÃO DE SISTEMAS DE ENSINO BRASILEIRO," em *XXXIX Congresso Brasileiro de Educação em Engenharia*, Blumenau - SC, 2011.

Support Vector Machine-based Method for Navigation of Mobile Robots through Paths in Plantations

Danilo Samuel Jodas[1, 2]

[1] CAPES Foundation, Ministry of Education of Brazil, Brasília – DF, 70040-020, Brazil
[2] Faculty of Engineering of the University of Porto. Rua Dr. Roberto Frias, s/n, 4200-465, Porto, Portugal
{Danilo Samuel Jodas, danilojodas@gmail.com}

Abstract. Nowadays mobile robots are used in several activities as exploration in areas of hazardous conditions, i.e., rescue operations and space missions, for instance. Besides, an effort to use mobile robots in agriculture for planting tasks and for keeping the application of pesticides within minimal amounts to mitigate the environmental pollution also has been done. In this paper a method to control the navigation of an autonomous mobile robot through tracks in plantations is presented. Track images are used to control robot direction by preprocessing them to extract image features. Such features are then submitted to a support vector machine in order to find out the most appropriate route. The goal of this work is to use a machine learning algorithm similar to artificial neural networks that decrease the training execution time and offers results next to those generated by the second one. In this work is also related the implementation of the support vector machine-based method in VHDL language to control the execution of the algorithm on a hardware platform. In this paper is reported the implementation of a support vector machine, which presented around 93% accuracy in predicting the appropriate route.

1 Introduction

In the recent years, there was a significant increase in the use of mobile robots in several areas. This increase is due to the execution of activities in places of difficult access or in situations that are potentially harmful to humans.

The appearance of intelligent algorithms was a factor that significantly contributed to the advancement of mobile robots due to the possibility of development of intelligent agents that autonomously and reliably perform the activities for which they were projected. Several papers on mobile robots report the use of artificial neural networks (ANNs) as the chosen intelligent system [1][2]. ANNs determine a result based on environmental input data captured by appropriate sensors. These results are used to control the mobile robot.

The combination of intelligent algorithms with image processing has provided good results for those applications in which the mapping of actions on the environment's plan is required to determine possible mobile robots' routes, avoiding collisions [3]. In such situations the images are captured by a video camera being submitted to preprocessing algorithms for the extraction of important image features, which are then input to the intelligent algorithm.

Mobile robotics benefits from more powerful processors reaching the market as several associated tasks require real time responses without which the implementation of the corresponding control algorithms is not viable. Hardware implementation of such algorithms is also an important issue as it allows for faster execution. Field Programmable Gate Array (FPGA) devices are frequently used due to the possibility of internal reconfiguration of logical devices, meaning that it is possible to adapt the hardware to specific algorithmic requirements, and eventually leading to an increase in the robot's performance.

The goal of this paper is to present a navigation system based on image processing algorithms and support vector machines. The navigation system is used to control the robot's direction, keeping it within the appropriate plantation track based on images of the terrain. Image processing algorithms are used to improve image quality as well as to extract relevant features from the image. The track thus identified is then processed by a support vector machine to determine the angle from the current direction of movement that the robot is required to deviate, if any. Besides, a comparison of both, SVM and ANN was performed to ascertain the one presenting the best outcome. The main goal of this comparison is to demonstrate that support vector machines can offer faster training and similar generalization capacity to the ANN for determining the steering angle. The support vector machine-based algorithm was implemented in VHDL (Very High Speed Integrated Circuits Hardware Description Language) as a circuit to directly control the steering angle into a hardware platform.

Navigation control is just one of many tasks performed by mobile robots. The correct navigation control is an essential task for the success of other applications related to autonomous robotics. The most important tasks are related to agriculture, such as weed detection and pesticide application in plantations, which contribute to reduce the environmental pollution. Therefore, the navigation system presented in this paper can leverage the use of mobile robots in agriculture by determining the most appropriate steering angle of a robot driving through a plantation based on the application of a Support Vector Machine as the machine learning algorithm on the information extracted from images of different plantation fields.

The remaining of this paper is organized as follows: in section 2 some relevant related works are discussed; in section 3 an overview of the system is presented; in section 4 test results for the recognition software are shown; and in section 5 some conclusions are drawn.

2 Related work

Several work in mobile robotics were already presented by various authors, mainly with respect to navigation and obstacle avoidance.

Mohanty et al [4] described in their paper a navigation system for mobile robots based in ANFIS (Adaptative Neural Fuzzy Inference System), which is a hybrid model that combines the features of both, Artificial Neural Network and Fuzzy Inference System. Information from mobile robot’s sensors such as left, right and front obstacle distance and the target angle are transmitted to an ANFIS, which is composed by five layers to determine the steering angle of the robot. The navigation system has two tasks: avoid the mobile robot collide into obstacles and move it towards target.

Sales et al [5] developed a vision-guided navigation system using Artificial Neural Network and Finite State Machine for identifying navigable or non-navigable image's block and determining the side that the mobile robot need turn. The algorithm presented in their paper divide the image in block of K x K size, from which are obtained a feature value that can be the average of both, RGB (Red, Green and Blue) and HSV (Hue, Saturation and Value) values, entropy and others information from block's pixels. The feature values from each block are transmitted to four or five artificial neural network, which determine whether the block is navigable or not. The block is navigable if the ANN' outcome is 1 and non-navigable otherwise. After processing each block, a navigable map with information of navigable and non-navigable regions of the image is generated. A Finite State Machine is used to control the driving of the mobile robot in navigable regions, as well as to keep in memory an action to be performed later, such as turn to the left when the robot need of a 90 degree turn left. The navigation system was embedded on a Surveyor SRV-1Q mobile robot.

Young-Jae Ryoo [6] presented a visual navigation system for mobile robot based on image processing and artificial neural network. The system receives as input eight points from the lines that represent the road's edge and these points are passed to the artificial neural network, which calculates the output that is the steering angle of mobile robot. The inputs are eight points that cross transversely with the edges of the road. The neural network was trained with the backpropagation algorithm.

Åstrand and Baerveldt [7] presented a mobile robot to detect weed in sugar beets. Detection of weed and path navigation is made based on images of terrain captured by video cameras. Steering angle is calculated by a modified line detection algorithm based on the Hough transform, which detects lines identifying the plant row. Åstrand and Baerveldt's mobile robot system does not use artificial intelligence to detect weed or to calculate the steering angle.

3 The navigation system

The navigation system is composed by two modules: the image processing module and the recognition module. The image processing module aims at improving image quality as well as extracting features from the plantation images. In the preprocessing submodule filters for image smoothing are implemented to minimize the noise present in the image. A segmentation submodule is used to separate the track from planted area. A skeletonization submodule is used to extract the corresponding central path. A translation submodule is used to move the identified path to the center of image in order to provide a standard data representation for the recognition module [8]. The recognition module is composed by a support vector machine, which receives the pixels of the translated skeleton and determines an output that represents the steering angle to the pattern of pixels presented. In Fig. 1 a diagram of the system operation is presented.

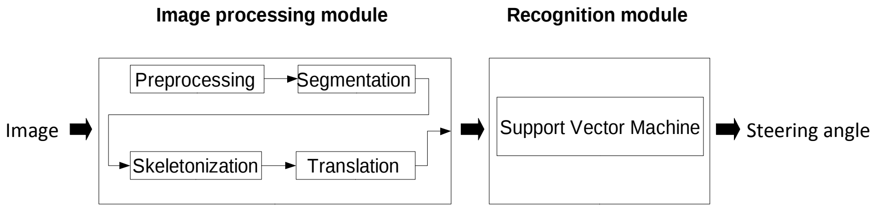


Fig. 1 - Diagram of system operation

The navigation system was developed in software for realization of initial tests. The development of image preprocessing algorithms was accomplished in C programming language together with the OpenCV library [9]. The development of support vector machine algorithms was also accomplished in C programming language.

3.1 Image processing module

For the initial tests an image bank composed by peanut and soybean plantation images was formed. The images were shot with a Kodak digital camera, M531 model, with a resolution of 6 Mega Pixels, under various illumination conditions. The bank has 1186 images, being 570 peanut images and 616 soybean images. Each image was resized to 300x225 pixels to reduce the preprocessing and segmentation algorithms' runtime. In Fig. 2 the processing steps of a plantation image are shown.

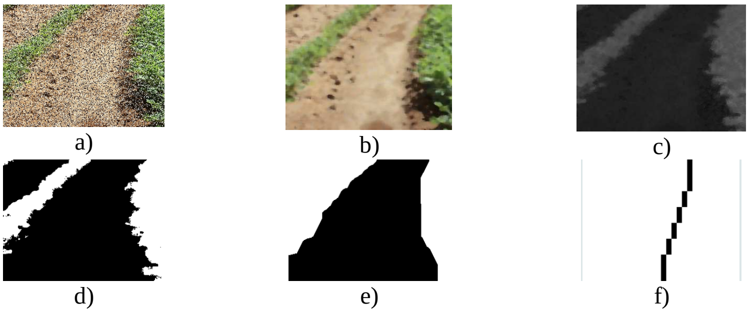


Fig. 2 - Steps of the image processing (a) Noisy image (b) Noise removed with median filter (c) Hue component of the image (d) Segmentation of path. (e) Edge smoothing and elimination of double path. (f) Skeletonized image

The image preprocessing is the step where techniques are applied to improve the quality of the image. The images of the bank were submitted to salt and pepper noises for testing the smoothing filters. The median filter was used to remove the noise present in the images, showing better results as compared to the low pass filter. The median filter was much efficient in removing noise from and preserving edges in the images. In Fig. 2(b) the result of the median filter applied with 7x7 template is presented.

The extraction of path's features consists in obtaining a binary image in which the path is characterized by the black color and the area corresponding to the plant is represented in white. The images of the bank are represented in the RGB color model and were converted to the HSI color model so as to obtain the hue, saturation and intensity channels. The HSI color model was chosen due to the possibility of color identification regardless of illumination intensity or degree of saturation of the color image. The advantage in use the HSI color model is the separation of the color information from intensity information. In Fig. 2(c) the hue component of an image is shown.

To smooth the border of the identified path, i. e., to remove small holes and links to adjacent paths, the morphological operation of closure was used. This procedure helps in the better representation of path's skeleton and minimizes the prominences effect in the image border. In Fig. 2(e) the result of application of the closure is presented. The structuring element has the shape of an ellipse with 20x20 pixels. 20 iterations of the operation of closure upon the segmented image were used to obtain a smooth enough border. The choice of structuring element and of iteration's count was made empirically based on various tests. As ellipse shaped elements present smoother borders than other structuring elements they leave the image border smoother too.

Due to the position of the camera with respect to the terrain it is possible that some images include paths adjacent to the main path. The elimination of those paths reduces information to be processed by the support vector machine. To remove such paths the region growth by pixels aggregation algorithm was used. The algorithm starts by randomly selecting seed pixels, which are represented by black pixels belonging to the path. The number of seed pixels is dependent on the amount of paths identified. For example, in the image of Fig. 2(d) two paths were identified, so it is possible to select two seed pixels, one for each path, and each path is processed as a region. Counting the pixels in each path is simultaneous with executing the region growth procedure. The regions are labeled with an integer value and in the end of the process it is possible to determine the region that has more pixels, which prevail in the image. The other paths are removed just altering the color for white. In Fig. 2(e) the result of the region growth algorithm execution is presented.

The next step is the thinning of the identified path. This is done by applying the thinning algorithm proposed by Zhang and Suen [10]. This algorithm reduces the number of path's information to be transmitted to the support vector machine. In Fig. 2(f) the result of the application of the thinning algorithm is shown.

The final step of the image processing module was resizing the image to the resolution of 30x23 pixels to reduce the size of the support vector machine input layer.

As the training of the SVM is supervised, the training parameters of each skeletonized image are needed for this procedure. So, the Hough transform [11] and the Least Square method [12] are applied to each image to obtain the approximate angle value to be used during the supervised training of the SVM.

3.2 Recognition module

Each image was transformed into a 690-element vector that corresponds to the image pixels. These values are inputs to support vector machine used for the recognition of steering angle. The images are classified as one of 19 possible patterns according to the angle they form to an imaginary straight line on the axis of movement. Such angles range from minus 45 degrees to plus 45 degrees with respect to the above mentioned straight line, and are discretized in 5 degrees intervals. As the values of the steering angles are discretized, the values produced by both, the Hough transform and the Least Square method are also discretized by using a nearest neighbor criterion, meaning that if the value θ of the Hough transform is 42, the two possible neighbors are 40 and 45. Thus, the desired output for the image will be 40 as this is the nearest value to 42.

Some SVM [13] kernel functions are shown in Table 1.

Table 1- Types of kernel functions

Type	Kernel function
Radial Basis Function	$\exp\left(\frac{-1}{2\sigma^2}\left\|x-y\right\|^2\right)$
Hardware friendly kernel	$2^{-\gamma\left\|x_i-x_j\right\|_1}$

The last kernel function presented in Table 1 is named hardware-friendly kernel (HFK) and was proposed by Anguita et al. [14] to reduce the implementation complexity of kernel functions in hardware. The use of this kernel is more suitable for hardware implementation than the Gaussian kernel because is not necessary to calculate divisions and exponentials, which makes the so called hardware-friendly kernel faster for several pattern recognition procedures. The parameter γ is an integer value defined as two to the power p (i.e. $\gamma = 2^p$, with p = 0, 1, 2 …) [14]. The L1-norm of the distance between the inputs and the support vector (i.e. ||x||) is represented as $||x_i - x_j||$. By using the norm instead of the Euclidean distance it is possible to avoid the computation of square root and exponentiation, which makes the hardware implementation less complex.

4 Experimental results

The tests were performed with 1128 images of peanut and soybean plantations. 71 images were used for training and 1057 images were used for verification. SVM and ANN were used for recognition of the steering angle.

The following information were considered in the tests:

- Amount of exact hits
- Amount of approximate hits
- Errors
- Overall hit percentage

The amount of exact hits indicates the number of images for which the required steering angle and the one computed by both, the SVM and the ANN exactly match. The amount of approximate hits indicates the number of images for which the difference between the required steering angle and the one computed by both, the SVM and the ANN is 5°. The errors represent the number of images for which the difference between the required steering angle and the one computed by both, the SVM and the ANN is above 5º. The overall hit percentage was computed as the ratio of the number of exact hits and the number of approximate hits. The amount of approximate hits was considered due to the similarity among the skeletons whose angles differ by 5º or less, causing both the SVM and the ANN to classify the corresponding image to an approximate angle figure. Some tolerance may be acceptable if the processing time is low enough to prevent the next exit to be made after a long distance can be travelled by the mobile robot, thus minimizing the off route displacement.

4.1 Test with SVM

In this test 71 training images and 1057 verification images were used, which belong to the plantation images' bank. Thus, the SVM's topology was represented by 690 input elements, 71 hidden elements and 1 output element. The output is a real value between -45 and 45 that represents the steering angle for the presented image. As kernel functions, the radial basis function and the hardware-friendly kernel were used. As discussed in the previous section an attempt was made to use a simpler kernel that is easier to implement in hardware, which also presents results close to those of the RBF kernel. The so named hardware-friendly kernel was tested, being these results compared to those of the RBF kernel. The value chosen for the parameter γ was 0.0625, which was determined empirically through tests showing the best results.
The results of this test are shown in the Table 2.

Table 2. Results of the test with SVM

Information	RBF	HFK
Exact hits	776	810
Approximate hits	197	175
Errors	84	72
Overall hit percentage	92.05%	93.19%

It is possible to notice that the results of the hardware-friendly kernel were better than those of the radial basis function kernel.

4.2 Test with ANN

Four tests with ANN has been performed. In each test different numbers of neurons in the hidden layer as well as different training parameters were used. In Table 3 the training parameters used in the tests are shown and in the Table 4 the results of the verification step are presented. It is possible to notice in Table 4 that the best result was provided by the third test. However, such result is inferior to the test performed with SVM. Besides, the amount of elements in the hidden layer of the ANN is greater than those used in the SVM topology. It is a disadvantageous characteristic because in this case twice as much hardware space would be needed to implement the ANN's topology.

Table 3. ANN's parameters in each test

Test	Learning rate	Momentum rate	Mean error	Number of hidden neurons
First test	0.7	0.1	0.0001	71
Second test	0.7	0.1	0.0001	115
Third test	0.7	0.1	0.00006	142
Fourth test	0.7	0.1	0.0001	213

Table 4. Results of the recognition using ANN

Information	First test	Second test	Third test	Fourth test
Exact hits	725	644	748	681
Approximate hits	215	252	206	233
Errors	117	161	103	143
Overall hit percentage	88.9%	84.8%	90.3%	86.5%

4.3 Performance analysis

The tests were performed on a computer with 4 Gigabytes of RAM memory and an Intel Dual Core processor of 2200 MHz. In Fig. 3 it is shown the execution time for each image processing algorithm.

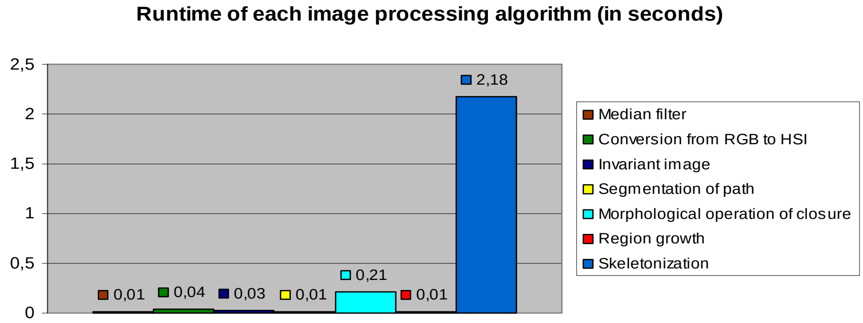


Fig. 3 - Runtime of image processing algorithms

It is possible to notice that the skeletonization algorithm consumes more execution time with respect to the other algorithms used, being this amount tenfold that required for the second largest execution time.

The execution performance for each SVM kernel and for ANN is presented in Fig. 4. As can be seen in the graphics, the runtime of the HFK was approximately 50% faster as compared to the RBF kernel in both training and execution. Besides, the HFK presented better results than the RBF kernel. It is possible to notice that the ANN training time took approximately 1 hour and train it several times until an ideal solution is reached could take hours or days, particularly if it is needed to include more training examples or change the ANN topology.

Independently of the kernel function used, the SVM proved more advantageous than ANN concerning to the training runtime because there are not backpropagation cycles in the former as there are in the latter.

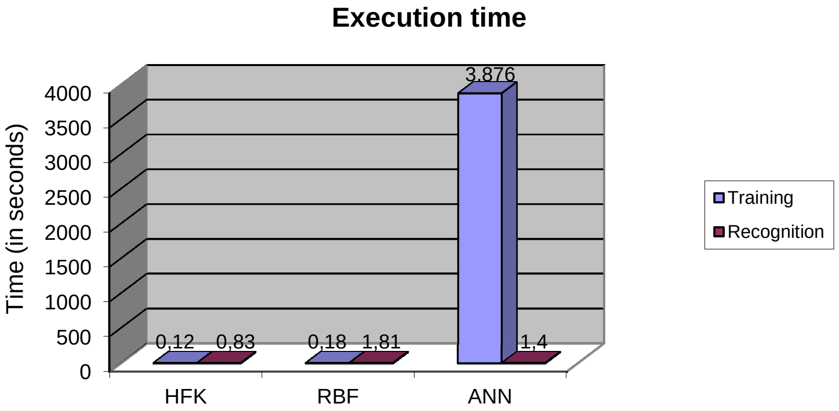


Fig. 4 – Execution time for training and execution

4.4 Implementation in VHDL

The SVM topology was translated in VHDL language [15] as a three-input and two-output circuit. The first input is reserved for the clock signal, the second one represents the reset signal and the last one is an 8-bit input that receives the pixels of the image. With respect to the outputs, the first one is a 1-bit signal used to indicate whether more pixels are necessary to be read or not and the second one is the steering angle recognized by the SVM circuit.

The execution steps were controlled by a Finite State Machine, which is a structure that controls the execution of a task by state transitions. A Finite State Machine is mathematically defined as M = {Q, I, f, qo, F}, where Q is a states set, I represent the inputs, f is a state-transition function, qo is the initial state and F is a stop condition state set [16]. A FSM starts the execution at an initial state qo and the transition through states is performed by the state-transition function f right up the stop-condition states. In digital circuits, the initial state is controlled by the reset signal and the state-transition function is implemented by a combinational circuit, which receive the actual state and the inputs signals to determine the next state. The result of the circuit is implemented by a combinational circuit which determine the output based in the actual state and, possibly, in the input signals.

In Fig. 5 it is shown the SVM state machine implemented in VHDL. The Hardware Friendly Kernel was used as the kernel function of our SVM because the results of this kernel were better compared to those from the RBF kernel. Besides, Hardware Friendly Kernel is less complex for the hardware implementation.

The description of each state is presented below:

- **init_0 and init_1**: represents the initial states
- **dist**: calculates the L1-norm distance between the input images and the training samples.
- **conversion**: performs the conversion to the IEEE 754 floating point standard [17] of both weights and L1-norm result values.

- **mult1, mult2** and **exponential**: calculate the HFK function presented in Table 1. The multiplication is done within 5 clock cycles and the exponential is done within 17 clock cycles. The same operation is done for mult3, sum and floattoint states.
- **mult3**: calculates the multiplication of the weight value by the HFK kernel result.
- **sum**: calculates the accumulated sum of the former state results. This operation is done within 7 clock cycles.
- **floattoint**: converts an IEEE 754 floating point standard value to an integer value. This conversion is done within 3 clock cycles.
- **round**: the value resulting of the former state is discretized using the nearest neighbor criterion discussed in Section 3.2. This operation is done within 3 clock cycles.
- **end**: represents the final state, which presents the steering angle to the circuit's output port.

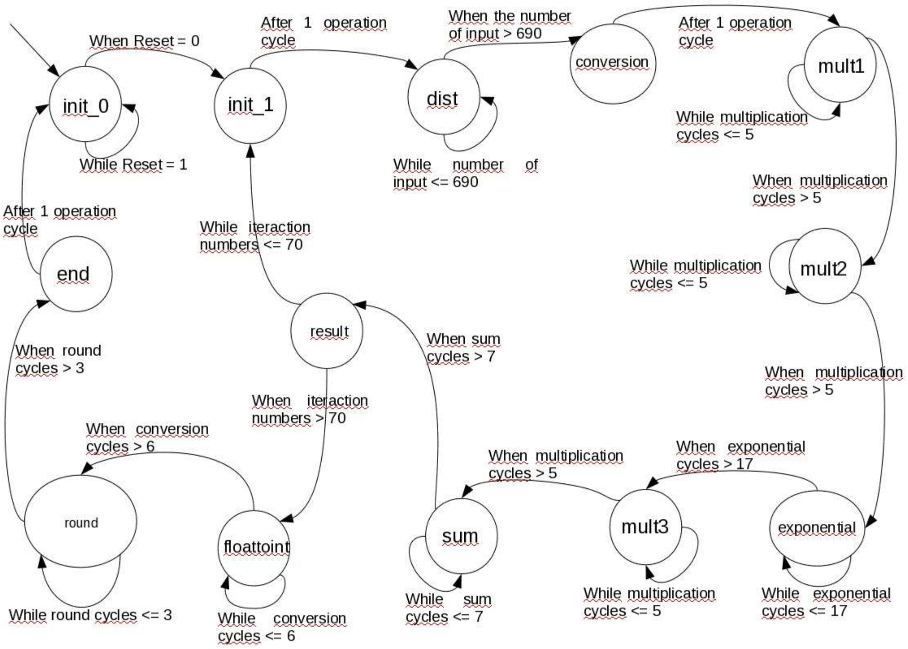
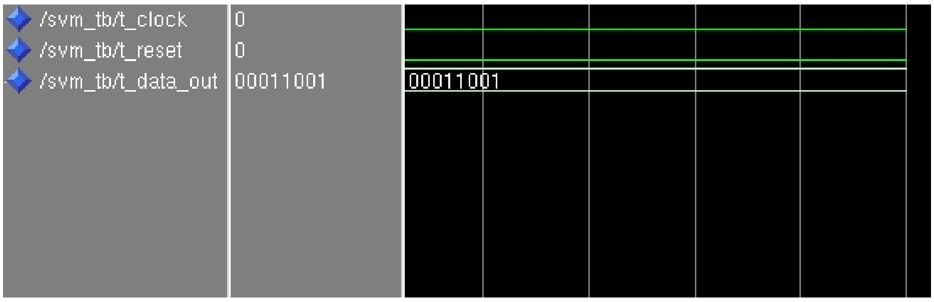

When Reset = 0
After 1 operation cycle
When the number of input > 690
After 1 operation cycle
init_0
init_1
dist
conversion
mult1
While Reset = 1
While number of input <= 690
While multiplication cycles <= 5
When multiplication cycles > 5
After 1 operation cycle
While iteraction numbers <= 70
end
While multiplication cycles <= 5
mult2
result
When sum cycles > 7
When round cycles > 3
When iteraction numbers > 70
When conversion cycles > 6
When multiplication cycles > 5
When exponential cycles > 17
When multiplication cycles > 5
round
floattoint
sum
mult3
exponential
While round cycles <= 3
While conversion cycles <= 6
While sum cycles <= 7
While multiplication cycles <= 5
While exponential cycles <= 17

Fig. 5 - State machine implemented in VHDL

A simulation using the ModelSim Starter Edition software [18] has been done in order to verify the results obtained from the SVM circuit before implementing it on the FPGA. The simulation took 20 nanoseconds, corresponding to a 50 MHz FPGA. In the simulations 20 segmented images were used for verification. The SVM circuit outputs were compared to those computed with the SVM algorithm implemented in the C programming language. The steering angles of all ModelSim simulations exactly matched to those computed by the C program. In Fig. 6 it is shown a simulation snapshot captured from ModelSim software.



```
/svm_tb/t_clock     0
/svm_tb/t_reset     0
/svm_tb/t_data_out  00011001   00011001
```

Fig. 6 - Snapshot from ModelSim software

According the outcome presented by the software the steering angle of this image is 25 degrees. Note that the result of the simulation matches with the one obtained from the C program. Thus, the VHDL implementation of the SVM circuit proved appropriate to be embedded into an FPGA device.

5 Conclusions

This work presented a navigation system for mobile robots handling paths in plantations, based on computer vision and support vector machines. The main objective of the navigation system is to obtain a satisfactory success rate in the prediction of steering angle using a small amount of image information, i. e., only the skeleton of the plantation pathways. Besides, was presented the comparison in the recognition of both SVM and ANN. The results obtained from SVM were better and the SVM's execution time was less than the ANN in training step. The ANN's training can to take hour or days as a SVM can perform it in just some minutes. It enables the execution faster of several training using different parameters and patterns to adjust the connection's weight and to achieve an appropriate recognition.

The major difficulty at this stage of the work was to achieve an exact percentage of correct recognition, a fact that is due to similarity of the skeletons obtained from those images having similar values of angles. However, for most images it has been verified that the difference between the angles computed by the SVM to that one computed using both the Hough transform and the Least Square method is about 5º. Analyzing the possible deviations from the expected route when such an error occurs was decided to accept this approximation as a correct guess. Nonetheless, to adopt such an approach the image processing algorithms execution time must be improved. Moreover, possible improvements on the algorithm execution time can lead to the update of the driving angle at smaller distance intervals.

The implementation of the SVM model was done in the VHDL language for evaluating the capacity of controlling the navigation directly into a hardware platform. The execution of the SVM circuit was simulated in the ModelSim software simulation and the results obtained matches with the ones provided from the software implemented in the C programming language. Thus, the circuit can be embedded into a hardware platform such as FPGA device for determining the steering angle. The advantage of executing the SVM into a FPGA device is the possibility of

reconfiguration of the logical devices to adapt to a specific algorithm, meaning that it is possible to increase the system performance with respect to the execution time.
A suggestion for future work is to implement a method for extracting some image features and use them as input of the SVM algorithm instead of using all image pixels. Besides, the implementation of the image processing module in VHDL language can be done in the future.

Acknowledgments

The author acknowledges CAPES – Brazilian Federal Agency for the Improvement of Higher Education – for the scholarship provided during this research.

References

1. Meng, M., Kak, A.C. Mobile Robot Navigation Using Neural Networks and Nonmetrical Environmental Models, IEEE Control Systems 13 (5) (1993), 30-39
2. Janglová, D. Neural Networks in Mobile Robot Motion, International Journal of Advanced Robotic Systems 1 (1) (2004), 15-23
3. Howard, H. Seraji. An Intelligent Terrain-Based Navigation System for Planetary Rovers, IEEE Robotics & Automation Magazine 8 (4) (2001), 9-17
4. Mohanty, P.K., Parhi, D.R., Jha, A.K., Pandey, A. Path planning of an autonomous mobile robot using adaptive network based fuzzy controller, in Proc. of IEEE 3rd International Advance Computing Conference (IACC) (2013), 651-656
5. Sales, D.O., Shinzato, P., Pessin, G.; Wolf, D., Osorio, F.S. Vision-Based Autonomous Navigation System Using ANN and FSM Control, Latin American Robotics Symposium and Intelligent Robotic Meeting (LARS) (2010), 85-90
6. Young-Jae Ryoo. Neural Network Control for Visual Guidance System of Mobile Robot, Lecture Notes in Computer Science 4432 (2007), 685-693
7. Åstrand, B., Baerveldt, A.-J. An Agricultural Mobile Robot with Vision-Based Perception for Mechanical Weed Control, Journal of Autonomous Robots 13 (1) (2002), 21-35
8. Yüceer, C., Oflazer, K. A rotation, scaling, and translation invariant pattern classification system, Journal of Pattern Recognition 26 (5) (1993), 687-710
9. Open Source Computer Vision, http://opencv.org, accessed in December, 2013
10. Zhang, T.Y., Suen, C.Y. A fast parallel algorithm for thinning digital patterns, Communications of the ACM 27 (3) (1984), 236-239
11. Gonzalez, R.C, Woods, R.E, Eddins, S.L. Digital Image Processing using MATLAB (2nd Ed.), Gatesmark Publishing, 2009
12. Bjõrck, A. Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, April 1996
13. Haykin, S. S. Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999
14. Anguita, D., Pischiutta, S., Ridella, S., Sterpi, D.. Feed-Forward Support Vector Machine Without Multiplier, IEEE Transactions on Neural Networks 17 (5) (2006), 1328-1331
15. Ashenden, P.J., Lewis, J. The Designer's Guide to VHDL (3rd Edition), Burlington: Elsevier, 2008
16. Grout, I. Digital Systems Design with FPGAs and CLPDs, Burlington: Elsevier, 2008
17. Kahan, W. IEEE Standard for Binary Floating-Point Arithmetic, http://grouper.ieee.org/groups/754/, accessed in December, 2013
18. ModelSim-Altera Software, http://www.altera.com/products/software/quartus-ii/modelsim/qts-modelsim-index.html, accessed in December, 2013

Sentiment Analysis on Twitter's Portuguese Language

Tiago Daniel Sá Cunha

Faculdade de Engenharia da Universidade do Porto, R. Dr. Roberto Frias, 4200-465
Porto, Portugal,
tiagodscunha@fe.up.pt

Abstract. Sentiment Analysis is a Natural Language Processing task which has been increasingly more explored in scientific research. Although there is already a large and matured application of the Sentiment Analysis methods for several languages, Portuguese is still one of the languages where this type of task is still being explored. In this paper, we aim to adapt State of the Art approaches on Sentiment Analysis to the Portuguese language. We will use as basis documents extracted from Twitter. These texts are previously filtered by language and also by subjectivity and objectivity. We apply several text pre-processing tasks and execute an automatic labelling procedure to train a classification algorithm. We experiment with several types of features and classification algorithms and conclude on which are the best pairs of feature sets and algorithms for our learning task. We conducted evaluations on 2 classes and 3 classes classification tasks based on the sentiment polarities (positive, negative and neutral). The best results were achieved using SentiLex and a combination of SentiLex and Bigrams as features using a Support Vector Machines (SVM) and Naive Bayes classifiers.

1 Introduction

Twitter[1] is a Social Media platform specialized in microblogging. It is one of the most used platforms for this purpose and allows access to large volumes of data that can be used for scientific researches, including Information Extraction tasks. Some examples can be found in platforms such as TWEETFEEL[2] and POPSTAR[3]. Data access is achieved through various Application Programming Interfaces (APIs), which enable access to the users profiles and their public messages. Furthermore, filtering by several aspects (including by content and by user), is also available through the Streaming API [1].

The scientific area of Natural Language Processing (NLP) is associated with several learning tasks which enable the extraction of knowledge from human-readable text [2]. Examples of these tasks are syntactic and grammatical error correction, entity recognition, term disambiguation, SPAM detection, authorship

[1] https://twitter.com/

[2] http://www.tweetfeel.com

[3] http://www.popstar.pt/

detection and sentiment analysis or opinion mining [2]. Although this area has its foundations in Artificial Intelligence and Linguistics, it soon expanded to several other areas, such as Information Retrieval, Machine Learning and Data Mining.

Sentiment Analysis (or Opinion Mining) is defined as the task of finding author's opinions in text regarding specific entities [3]. This scientific area has several tasks, such as creation of sentiment lexicons, opinion summarization, opinion search and extraction and spam detection in opinion texts.

Microblogging is a popular choice for Sentiment Analysis mining due to: 1) the existence of users with different characteristics that express their opinions regarding several topics, 2) the high scale and easy accessibility to texts and 3) Twitter's audience is characterized by multiple nationalities and with growing range [4].

Sentiment Analysis is a on-going research topic, although in some languages its results are better, due to extensive research. In the Portuguese language there is a growing interest in this topic, associated with the vast usage of Social Media platforms and the subject is far from being perfectly resolved. Furthermore, there is interest in developing Sentiment Analysis strategies for Social Media lexicons, since there are many advantages to extract from it, including access to a large number of individuals who express sentiments and focus on specific users to evaluate their actions both in the virtual and the real worlds.

This paper aims, therefore, to adapt State of the Art approaches for Sentiment Analysis in the English language to the Portuguese language. It is expected that the results follow the same behaviour as those developed for other languages, if the pre-processing tasks are able to reduce the problem to a similar structure. Another expected result is that, in the same manner as in the State of the Art approaches, the classification algorithms that yield better results are Naive Bayes and Support Vector Machine.

The goals set for this paper are: 1) extract tweet messages from Twitter and filter them by language, objectivity and subjectivity to build a text lexicon for Sentiment Analysis; 2) apply automatic labelling procedures on the dataset for the classification algorithm; 3) experiment with different feature sets and classification algorithms.

This paper is divided in 5 section: the Related Work (section 2) where we list the different types of Sentiment Analysis methodologies in the State of the Art, the section 3 that describes our approach for Sentiment Analysis, including data extraction and preparation, data analysis, data labelling, feature extraction and classification, the Results of our experimental setup (section 4) and the Conclusions of our methodology and Future Work (section 5).

2 Related Work

Sentiment Analysis is a hot research topic nowadays and, therefore there are several State of the Art compilations, such as those found in Pang [5] and Ling [6]. However, since there is work in Sentiment Analysis especially developed for Twitter, it is important to analyse in depth these types of approaches.

Sentiment Analysis is largely studied on Twitter data, by the usage of several strategies and pre-processing techniques [7–11].

For instance, Bravo-Marquez [12] aimed to discover which are the most effective metafeatures and techniques for supervised Sentiment Analysis approaches, through the most popular resources available. The different types of labelling classes are summarized in their work, namely: 1) polarity - classification is executed through a generic categorical variable (positive, negative and neutral), 2) emotion - the result of the classification is also a categorical variable, but related to an emotion term (for instance happiness and sadness) and 3) strength - the classification assumes a numerical scale based either on polarity or emotion. The approach consists in modelling a tweet as a vector of sentiment features extracted from sentiment datasets (OpinionFinder Lexicon, AFINN Lexicon, SentiWordNet Lexicon, SentiStrength, Sentiment140 and NRC), which possess all the different types of labelling classes for Sentiment Analysis presented previously. In their experimental work, the features were evaluated and combined to obtain the best result possible, using popular classification algorithms for Sentiment Analysis, such as Decision Trees (CART and J48), Naïve Bayes and SVM. Their conclusions state that 1) feature combination yields better results than independent features in State-of-the-Art approaches, 2) polarity-based features are the most informative when used independently in opposition to emotion-based, which provide the worst results, 3) most features are better to classify sentiment, rather than sentence subjectivity and 4) the best performance results for precision and F1 measures are achieved using a SVM classifier.

The supervised approach developed by Ghiassi [13] is specially oriented for sentiment analysis on a specific Twitter entity and uses a typical Twitter lexicon built by the authors. The lexicon on which the system is based is created via a N-Gram model with a statistical validation to build the features. The pre-processing tasks executed prior to to achieving the final model were stop-words removal, stemming, transformation of data in a vector space model, term weighting, feature selection and creation of training and testing datasets. Feature engineering techniques are applied to keep the most important features in the system, reduce the problem dimensionality and contribute to data noise reduction. Afterwards, the filtered model is compared with a dataset of manually annotated terms, accordingly to its sentiment polarity. The scoring model in this work is simultaneously categorical and numerical, with the following relationship: *Strongly Positive*(2), *Mildly Positive*(1), *Neutral*(0), *Mildly Negative*(-1) and *Strongly Negative*(-2). The classification step of the procedure is achieved using Neural Networks (DAN2) and SVM with a one-vs-all strategy. This means that 4 models were created for the all classes except the neutral one. The classifiers were assembled in a chain and those which reached the end of the chain without any classification, were regarded as belonging to the neutral class. The results for this work conclude that the better results were achieved with the Neural Network algorithm.

The research developed by Pak [4] aimed to create a supervised approach to classify tweets accordingly to its sentiment, having for basis a set of collected tweets. The main contribution of this work is an automatic process to label tweets accordingly to sentiment expressed by emoticons to be used directly in a classification algorithm. The existence of a positive emoticon in a tweet leads to its annotation as a positive message and the same happens vice-versa. For the neutral class, a set of news tweets are extracted, since it is assumed that they possess only objective sentences and, therefore, do not state any sentiment. Several pre-processing tasks are performed to clean the data, such as removal of URL, usernames (also known as mentions), special words (”RT” and hashtags) and emoticons. Features are engineered by N-Gram models (unigrams, bigrams and trigrams) and by POS (Part-of-Speech) tags. The classic N-Gram models are adapted to include the special negative term ”not”, since it is a sentence modifier and a special case in sentiment analysis. The classification algorithms tested were Naïve Bayes, SVM and Conditional Random Fields (CRF). To improve performance, features are filtered by two criteria: 1) by low entropy, which indicates that the feature has high importance to distinguish texts and 2) by ”salience”, a proposed metric to filter N-Grams that do not present sufficient discrimination in the task of Sentiment Analysis. The conclusions of this research are: 1) the best accuracy results are achieved using bigrams as features, 2) salience is the criteria that improves the most the performance in feature filtering, 3) POS-tags present highly accurate features to distinguish between subjective and objective texts and 4) the best results are achieved using a Naïve Bayes classifier.

Souza [14] developed an approach for Sentiment Analysis for Twitter in the Portuguese language. The work evaluates several types of pre-processing tasks, Portuguese polarity lexicons (OpLexicon and SentiLex) and a negation model specially developed for the Portuguese language. This negation model considers the words ”não”, ”nunca” and ”ninguém” as sentiment polarity shifting terms in a text. The strategy used to label tweets by polarity is the presence of hashtags ”#win” and ”#fail” for positive and negative messages, correspondingly. The results of this work show that: 1) the pre-processing techniques had low effect on the performance, 2) the best results were achieved using OpLexicon and 3) the models of negation introduced improved the classification precision.

3 Sentiment Analysis on Twitter Portuguese Language

The methodology we developed for Sentiment Analysis is based on adaptation of other State of the Art approaches for the Portuguese Language. We extract subjective and objective tweets to use an automatic Sentiment Analysis labelling procedure. Several data cleaning tasks are performed, based on State of the Art approaches and the impact of a set of linguistic features with several classification algorithms were evaluated.

3.1 Data Extraction and Preparation

We start by collecting a set of 300,000 tweets based on the WeFeelFine [15] methodology. We extract subjective tweets via Streaming API with a filtration accordingly to the presence of conjugations of the verb "sentir" ("feel" in English). We used conjugations for all personal pronouns in the Past, Present and Future times of the Indicative mode. This strategy is a sufficient condition to assert the existence of subjectivity in texts and enable the extraction of sentiments in these texts. However, we understand that these types of sentences are not the only ones that express sentiment, but rather a simplistic approach that fulfilled the goals of subjectivity.

To ensure the language filtering for the Portuguese language, we recurred to an external API. We used Xerox Language Identifier API[4] in a sub-process and stored only the texts that verify simultaneously the conditions of subjectivity and language.

Data is cleaned accordingly to the strategies found in the literature [4]. We filter undesired items from the data (URLs, RT, mentions (@), hashtags(#), punctuation, accents, invalid chars and emoticons) and perform tokenization to transform text into feature-based vectorial space.

3.2 Data Analysis

Prior to developing the Sentiment Analysis task, we performed some data analysis tasks. Term frequency and N-Gram models were analysed and the results presented below. In figure 1 we present the term frequency distribution of the top 50 terms.

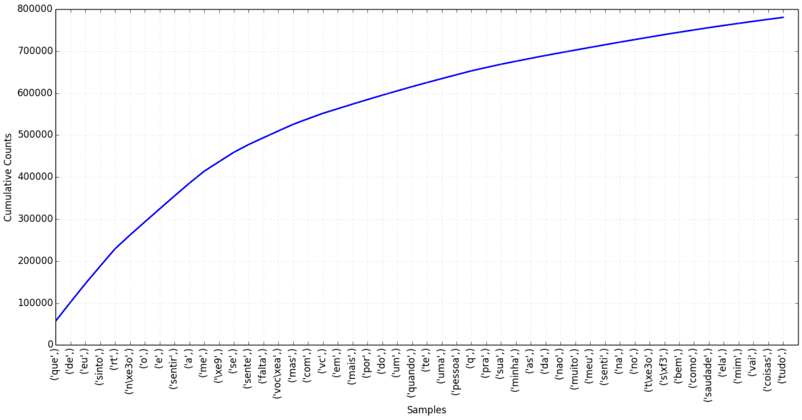


Fig. 1. Term frequency distribution for the top 50 terms.

The figure shows expected results, since among the terms we find mostly Portuguese stop-words and some conjugations of the verb "sentir" (feel).

[4] http://open.xerox.com/Services/LanguageIdentifier

Another common strategy to analyse data is to create N-Gram models, which allow to predict the next term in a sequence, based on previous terms [16]. The value N can assume any natural number and is translated to the number of consecutive terms considered for the model. We present the top 10 most frequent items in the 2-Gram (Bigram) model in Table 1. The models generated allows to verify that the most frequent Bigrams are essentially composed by conjugations of the verb ”sentir” and also stop-words. With these analysis, we concluded that the pre-processing tasks described previously had to be executed, in order to achieve term uniformity and disambiguation.

Table 1: Summary of frequent Bigrams.

Bigram	Frequency
('eu', 'sinto')	10470
('o', 'que')	9644
('que', 'eu')	8207
('me', 'sinto')	8100
('eu', 'não')	5324
('se', 'sentir')	4894
('se', 'sente')	4634
('que', 'sinto')	3946
('a', 'pessoa')	3711
('sinto', 'falta')	3480

3.3 Data Labeling

In order to build a labelled training dataset, a decision was taken to use the automatic approach using emoticons presented in the literature [4, 13, 10]. We consider the emoticons “:-)”, “:)”, “=)”, “:D” as positive and “:-(”, “:(”, “=(”, “;(” as negatives.

Therefore, we filtered the collected dataset by positive and negative emoticons, obtaining correspondingly 1,722 and 3,552 tweets for each class. This approach was chosen since it is an easily reproducible methodology that reduces the overhead of manually labelling the examples.

The neutral class was built using the strategy in Pak [4]. We selected a list of Portuguese news entities in Twitter and extracted 1133 tweets that are considered as objective, and therefore, that do not carry any sentiment whatsoever. The news entities we selected were ”Jornal de Notícias” (”@JornalNoticias”), ”SIC Notícias” (”@SICNoticias”), ”Diário de Notícias” (”@dntwit”), ”Público” (”@Publico”), ”Agência Lusa” (”@Lusa_noticias”) and ”Antena 1” (”@antena1rtp”).

Instead of removing some examples to balance the data, we decided that it would affect the normal behaviour of this strategy and would introduce bias with negative effect. Another strategy planned to fight this problem was to use

SentiLex's contents to filter tweets and ensure that each message would not contain words of another polarity and that it would contain at least one from the desired polarity. However, this strategy reduced drastically the number of examples to 258 and 1,472 for positive and negative classes respectively, unbalancing even more the data and introducing bias that would prevent the classifier from learning unexpected patterns.

3.4 Feature Extraction

The next step to obtain polarity annotated datasets is to choose the type of features to consider for the classification step. We decided to use N-Gram models (Unigrams and Bigrams), Bigrams with a negation awareness, SentiLex as features and combinations of the previous elements. We decided to execute, at first, classification for 2 classes only (positive and negative) and afterwards, select the most promising feature sets to evaluate the entire 3 class problem. Summaries of each feature extraction procedure can be found in Table 2. An incidence is considered as a table cell marked as True.

N-Gram models were created by joining the texts in all labelled tweets and stored independently to perform feature extraction posteriorly. However, this strategy leads to the creation of feature sets of high dimensionality in the Bigram case, as we can observe in Table 2. We chose to select the top 10,000 most frequent features to solve the problem, while reducing the matrix sparsity.

Table 2: Summary of feature extraction for the 2 and 3 class problems.

Feature type	Number of features	Number of incidences
2 Class problem		
Unigrams (U)	6666	49188
Bigrams (B)	32399	52650
Bigrams filtered (BF)	10000	52650
Bigrams filtered w/ negation (BFN)	10000	27421
SentiLex (S)	41590	6410
SentiLex filtered (SF)	1053	6410
Unigrams + Bigrams filtered (U+BF)	16666	71291
SentiLex filtered + Bigrams filtered (SF+BF)	11053	17219
SentiLex filtered + Bigrams filtered w/ negation (SF+BFN)	11053	33846
3 Class problem		
Bigrams filtered w/ negation (BFN)	10000	26121
SentiLex filtered (SF)	1300	7351
SentiLex filtered + Bigrams filtered w/ negation (SF+BFN)	11300	33488

The strategy for Bigrams with negation awareness is similar to the one found in Pak [4], although instead of using the word "not", we use its equivalent in

Portuguese ”não”. Therefore, if we have the sentence ”Eu não gosto de ti”, the Bigrams yielded are (”eu+não”,”gosto”), (”não+gosto”,”de”) and (”de”,”ti”). Souza [14] also implemented 2 strategies for negation, however we chose not to use it since their strategy is directly developed for Brazilian Portuguese.

Another feature set is to use SentiLex as features for Sentiment Analysis. We consider that this annotated dataset of polarized words, vastly covers simple words that are sentiment shifters, and therefore, useful to distinguish texts of different polarity. We also filtered this feature set, since many of the features were not matched by any tweet and were, therefore, discarded.

We decided to combine features attempting to improve the performance of the classifier, by increasing the number and type of elements to be included in the features. We combined Unigrams and Bigrams in a feature set, Bigrams and SentiLex in another and lastly Bigrams with negation awareness and SentiLex.

The extraction method is based on a bag-of-words model, where a verified incidence of a feature in a message was marked as True or marked as False if that matching was not present. The matrix in the format $message \times feature$ is then submitted for classification.

3.5 Classification

The classification algorithms selected to solve this learning task were RapidMiner[5] implementations of Naïve Bayes, Decision Tree, SVM and kNN. The algorithms were selected since most State of the Art approaches choose Naïve Bayes and SVM, while kNN and Decision Trees can also be found in some examples.

We experimented with each combination of feature set and classification algorithm in 2 types of problems: 2 class and 3 class Sentiment Analysis classification. The evaluation was executed using 10-fold cross-validation.

Due to its characteristics, SVM and kNN classifiers required the transformation of the matrices from True/False incidences to 1/0 numerical instances. Also, SVM is unable to perform classification for more than 2 classes by itself and required the setup to include a one-vs-all strategy. This strategy builds binary SVM classifiers for each class and joins the classification results after.

4 Results and Discussion

We present now in the Tables 3 and 4 the detailed evaluation results with the metrics average accuracy, precision and recall for each class implemented. Each Table highlights the best results for each feature set, based on the accuracy measure. The test dataset is composed by all the labelled examples: 1,722 positive examples, 3,552 negative and 1,133 neutral.

In the 2 class problem, the best performance was achieved with SF and SF+BFN using the algorithm SVM with an accuracy of 68.05%. However, the

[5] http://rapidminer.com/

changes in performance are minor considering all algorithms except kNN, that consistently yields the worst results. In fact, we observe that the algorithms Naïve Bayes, SVM and Decision Tree are the best choices, depending on the feature set. The more consistent algorithm in terms of performance is Naïve Bayes, with a balanced performances for accuracy, precision and recall.

Table 3: Experimental evaluation for the 2 class problem.

Feature	Algorithm	Accuracy	Precision	Precision	Recall	Recall
			positive	negative	positive	negative
U	Naïve Bayes	66.40%	48.49%	74.57%	46.52%	76.04%
	Decision Tree	67.12%	43.48%	67.54%	2.32%	98.54%
	SVM	67.22%	48.47%	68.07%	6.45%	96.68%
	kNN	58.51%	39.62%	72.51%	51.63%	61.85%
BF	Naïve Bayes	59.00%	39.56%	71.95%	48.43%	64.13%
	Decision Tree	67.48%	62.07%	67.51%	1.05%	99.69%
	SVM	66.84%	29.23%	67.31%	1.10%	98.70%
	kNN	35.06%	33.03%	74.90%	96.28%	5.38%
BFN	Naïve Bayes	65.02%	45.89%	72.60%	39.90%	77.20%
	Decision Tree	67.96%	64.81%	68.02%	4.07%	98.93%
	SVM	67.75%	56.60%	68.09%	5.23%	98.06%
	kNN	48.67%	36.59%	76.39%	78.05%	34.43%
SF	Naïve Bayes	67.61%	50.77%	71.12%	26.77%	87.42%
	Decision Tree	67.56%	53.90%	67.93%	4.41%	98.17%
	SVM	68.05%	54.57%	69.17%	12.83%	94.82%
	kNN	38.23%	33.26%	71.43%	88.62%	13.80%
U+BF	Naïve Bayes	65.59%	47.34%	74.89%	49.13%	73.56%
	Decision Tree	67.10%	43.30%	67.55%	2.44%	98.45%
	SVM	66.97%	38.10%	67.44%	1.86%	98.54%
	kNN	59.91%	40.00%	71.70%	45.53%	66.89%
SF+BF	Naïve Bayes	63.18%	44.31%	74.10%	49.77%	69.68%
	Decision Tree	67.58%	58.82%	67.69%	2.32%	99.21%
	SVM	66.95%	36.36%	67.40%	1.63%	98.62%
	kNN	40.50%	34.22%	76.34%	89.20%	16.89%
SF+BFN	Naïve Bayes	66.34%	48.30%	73.94%	43.84%	77.25%
	Decision Tree	67.88%	61.48%	68.03%	4.36%	98.68%
	SVM	68.05%	59.30%	68.39%	6.85%	97.72%
	kNN	54.44%	39.11%	76.76%	71.02%	46.40%

It is also observable that the unbalanced dataset consistently provides worse performance for the positive class, especially by analysing the recall measure for this class. This may indicate that the labelling procedure is not sufficient to distinguish the positive and negative classes, either by the reduced number of examples or by the inconsistencies in the texts. On the other hand, kNN shows the opposite phenomena, by obtaining a high recall for the positive class recall, although maintaining worse performance in the positive precision.

The 3 class problem yields the best performance results with the features

Table 4: Experimental evaluation for the 3 class problem.

Feature	Alg.	Accuracy	Precision	Precision	Precision	Recall	Recall	Recall
			positive	negative	neutral	positive	negative	neutral
SF	NB	59.46%	45.92%	61.78%	61.34%	24.51%	82.97%	38.89%
	DT	57.04%	68.75%	56.38%	92.79%	0.64%	99.69%	9.08%
	SVM	56.65%	43.72%	57.22%	68.09%	12.14%	91.39%	15.43%
	kNN	32.65%	27.13%	65.95%	72.99%	86.99%	13.91%	8.82%
BFN	NB	64.71%	41.21%	71.36%	74.05%	36.64%	71.11%	87.30%
	DT	57.66%	59.56%	56.81%	93.20%	1.86%	99.27%	12.08%
	SVM	46.04%	34.03%	51.13%	11.32%	10.45%	76.10%	5.91%
	kNN	41.22%	27.77%	73.50%	67.13%	72.47%	35.14%	12.79%
SF+BFN	NB	65.53%	43.98%	71.01%	78.93%	41.75%	72.55%	79.63%
	DT	57.74%	100.00%	56.75%	97.97%	0.35%	99.92%	12.79%
	SVM	53.30%	44.47%	55.39%	35.47%	12.60%	85.98%	12.70%
	kNN	44.77%	27.93%	74.05%	59.68%	64.40%	44.34%	16.31%

SF+BFN using a Naïve Bayes classifier and obtaining an accuracy of 65.53%. This experiment included only the features with highest accuracy and precision for the 2 class problem. We justify this selection since we consider these measures to be the most important for this type of task and also due to their highest performance rates.

In the 3 class problem, the difference of performance is more significant when we evaluate different feature sets. It is also observed that the overall performance for 3 classes decreased when comparing with the 2 class problem. The algorithm Naïve Bayes obtains the best results for all combinations of feature sets.

In this experiment, the precision for the negative and neutral classes is generally high across all combinations of feature sets and algorithms. The positive class is, similarly to the 2 class problem, the one with worse results, except on the case of SF+BFN with Decision Tree. This combination obtains precision of 100% for the positive class, although with extremely low recall of only 0.35%.

With these results, we can prove the hypothesis that it is possible to recreate a State of the Art approach to adapt a Sentiment Analysis methodology to the Portuguese language, as it was claimed by Pak [4], when they stated that their approach was language independent.

On the other hand, we are unable to prove that Naïve Bayes and SVM yield the best results in this type of approach for all different types of features. In some cases, Decision Trees exhibited the better performance results.

Although our approach covered several aspects that respected State of the Art researches and in fact we have reached a fully functional Sentiment Analysis approach for Portuguese language, our methodology may have some aspects to improve and is far from perfect. For once, stemming, stop-word removal and other text cleaning tasks can contribute to improve the performance of a classifier. On the other hand, there are some State of the Art features we did not implement or evaluated. Good examples can be found in the work developed by Mohammad [17]. Lastly, the negation models in Souza [14] could be included in our negation aware Bigram strategy and its performance evaluated.

5 Conclusions and Future Work

We developed a strategy for Sentiment Analysis in Twitter Portuguese texts, by adapting State of the Art methodologies. We labelled data automatically using emoticons to define positive and negative messages and considered that news tweets would define objective and sentiment-free texts. An experimental study was conducted with different types of features, including simple and negation aware N-Grams, SentiLex and feature combination. The classifications algorithms tested for each feature set were Naïve Bayes, Decision Tree, SVM and kNN. The results show that for a task to distinguish positive and negative messages, the best features are SentiLex or a combination of SentiLex and negation aware Bigrams using a SVM classifier. However, when the neutral class is included, the best results can be found with the combination of SentiLex and negation aware Bigrams as features for a Naïve Bayes classifier. However, since we executed an experimental stochastic procedure and the final evaluation result is averaged and with a standard deviation, we are unable to draw any conclusions about which combinations of feature sets and classification algorithms are best. Some tasks for future work range across all steps of the procedure: 1) devise a strategy to include hashtags and emoticons for data labelling, 2) perform stemming and stop-word removal for the Portuguese language, 3) include features studied in State of the Art, such as POS tags, 4) use salience and entropy to filter features [4] and 5) extend the number and type of classification algorithms.

References

1. Twitter. The Streaming APIs. Available at `https://dev.twitter.com/docs/streaming-apis`, 2012.
2. Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551, 2011.
3. Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.
4. Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010. European Language Resources Association (ELRA).
5. Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, 2008.
6. Bing Liu. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.
7. Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

8. Alec Go, Richa Bhayani, and Lei Huang. Twitter Sentiment Classification using Distant Supervision. *Processing*, pages 1–6, 2009.
9. Andrius Mudinas, Dell Zhang, and Mark Levene. Combining lexicon and learning based approaches for concept-level sentiment analysis. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, pages 5:1—-5:8, New York, NY, USA, 2012. ACM.
10. Arturo Montejo-Ráez, Eugenio Martínez-Cámara, M Teresa Martín-Valdivia, and L Alfonso Ureña López. Ranked WordNet graph for Sentiment Polarity Classification in Twitter. *Comput. Speech Lang.*, 28(1):93–107, 2014.
11. Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on Web search and data mining*, WSDM '13, pages 537–546, New York, NY, USA, 2013. ACM.
12. Felipe Bravo-Marquez, Marcelo Mendoza, and Barbara Poblete. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '13, pages 2:1—-2:9, New York, NY, USA, 2013. ACM.
13. M Ghiassi, J Skinner, and D Zimbra. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst. Appl.*, 40(16):6266–6282, 2013.
14. Marlo Souza and Renata Vieira. Sentiment Analysis on Twitter Data for Portuguese Language. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*, PROPOR'12, pages 241–247, Berlin, Heidelberg, 2012. Springer-Verlag.
15. Sepandar D Kamvar and Jonathan Harris. We feel fine and searching the emotional web. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 117–126, New York, NY, USA, 2011. ACM.
16. Daniel Jurafsky and James H Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 2000.
17. Saif M Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. *CoRR*, abs/1308.6242, 2013.

SESSION 2

INFORMATION SYSTEMS & COMMUNICATION NETWORKS

SOA/REST and Open Source as a practical approach for small business application development

Vasco Miranda

A Methodology Proposal for Reliability and Availability on WirelessHART: Best Practices

Mara Martins

Improving wireless sensor network efficiency in dense cluster

Benedito Bitencort

SOA/REST and Open Source as a practical approach for small business application development

Vasco Nuno Barreiro Capitão Miranda

Department of Informatics Engineering, Faculty of Engineering,
University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto,
Portugal,
prodei1300769@fe.up.pt

Abstract. Small businesses have largely been left out of the SOA paradigm. Motivated by Portugal's SME, which mostly consists of small businesses, this paper presents a proposal for a SOA implementation approach using Open Source tools in daily life development tasks. This paper tries to demystify the idea that it requires complex and costly development models and also aims to contribute to narrowing the gap between business and technical processes in order to unify business semantics. In this, it is proposed a practical REST approach in order to create an overall strategic plan and focus how architectural context can support a sustained transformation of functionality into services, as a facilitator of the challenges emerging in a migration approach of a cloud integration strategy. Starting from a real life scenario, it is presented a small prototype as proof of concept and is possible to verify that an SOA approach, using Open Source tools, is perfectly within the reach of small businesses as an added value.

Keywords: Service Oriented Architecture, SOAP, REST, Web Services, SaaS, SME, Open Source

1 Introduction and motivation

The current technical challenges that SME faces today require an assertive and adaptable dynamic on the market approach.

According the "SBA Fact Sheet 2012 – Portugal, DG Enterprise and Industry - Small Business Act Fact Sheet – 2012, EC - European Commission - Directorate General Enterprise and Industry", Portugal's SME sector is characterized by the large share of micro-firms with less than 10 employees. With 94%, it is two percentage points higher than the EU average, within a total of 99,9%, having 5% of small and 0,7% of medium sized enterprises. Only 0,1% are large enterprises.

This article is motivated by those facts and by the knowledge of the functionality based traditional model in software development, proposing a new approach.

The emergence of scalable, multi-platform web services and their proliferation, together with the heterogeneity of existing solutions implies a rethinking of strategies for information systems, from a sustainable logic point of view. The adoption of SOA in business computing environments is growing due to the significant of cost reduction, in the planning, deployment and operation of IT projects. However, the organic transformation from legacy enterprise applications to SOA applications has occurred mostly in large enterprises. Small businesses have largely been left out of the SOA transformation. For small businesses and organizations, SOA is a mysterious technology little related with their business. (Castro-Leon E. & He J. & Chang M., 2007)

The main goal of this paper is to demonstrate a scenario where the using of SOA is possible with open source resources and not to heavy work, making possible for small business to enjoy the benefits of service oriented architecture by changing the software development paradigm approach.

It was written in order to give the reader some basic and simple concepts that can contextualize, justify and provide a better understanding of the main goal.

From a basic question - is achievable for small business, without the use of complex mechanisms and proprietary solutions, change the development paradigm solutions in a SOA approach, benefiting of its advantages, using Open Source platforms? - It is intended to demonstrate, by experimentation and possible reproduction, an approach that respects SOA principles, featuring the involved technologies and implementing a small prototype as proof of concept. A real life scenario is used, where the prototype acts for a starting point for a future product.

The context of this article is SME's reality. For that reason, the proposed approach focus on a small businesses logic and their environment, in terms of size and IT human resources, typically few, and in some cases, none, one or two elements.

Business strategies increasingly require solutions that cover a wide range of users, departments and business processes. However, the effort required to address the heterogeneity of solutions leads to redundant and impractical maintenance structures, regarding costs and technical complexity.

On an ideal scenario, the principle of "develop once, deploy everywhere" should always be present in any strategy of information systems development. Moreover, the large number of software solutions that contribute to the various business processes will certainly involve, at some stage of the company growth, systems integration mechanisms. SOA approach can have vital role as integration middleware, not only on a logical purpose, but also as a semantic transformation of the technical processes into services.

One of the major problems in the daily lifes of small businesses is the gap between IT and Business processes, and the difficulty of orchestrating them in a common language and goals, or, in other words, regardless of technological processes, is important to ensure the presence of a business logic that is not always in the knowledge of IT departments. Simplistically, the goal should not be on the functionality or technical processes, but rather in the services, as they provide, in order to narrowing the gap between business processes and IT departments.

Another challenge that companies face is the no evolution of the IT human resources skills for new development methodologies, being stuck to traditional models, unorganized and with difficult interpretation by external elements, therefore, resistant and fearful of the complexity new approaches will bring.

Thus, in addition to bringing the business and IT processes in a language and a common commitment, is also a motivation of this article, contributing to demystify the idea that SOA approaches require complex models of development.

This paper is structured as follows: After an introduction to the topic and explanation of motivation in context (Section 1), some features are presented of the proposed conceptual approach (SOA). A reference is made to the strong link between the proposed approach and its availability on the cloud, and then is presented the characteristics of the two major forms of implementation (Section 2). On this assumption, after a contextualization and the chosen approach justification (Section 3), a small demonstration prototype implementation is shown. Finally, in Section 4, some points of discussion are presented as well as some conclusion and future work.

2 BACKGROUND AND TECHNICAL APPROACH

2.1 Service Oriented Architecture

SOA (Service oriented) is an architecture for building business applications by means of services. Is not a new concept, it's an evolution from the traditionally tightly coupled models. Can be described as a natural evolution of distributed systems, which allow components from different systems connect through services. SOA is now a standard and a reference model for service oriented computing, as a supporting architecture of Web Services.

The OASIS SOA Reference Model defines SOA as: "Service Oriented Architecture (SOA) is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains".

According to (Ashish S. et al., 2012) "An SOA is a component model that interrelates services (the different functional units of an application or an enterprise) through well-defined interfaces and protocols between these services". Although existing systems can not simply be replaced, an SOA approach can bring agility and flexibility to existing systems, bringing opportunities for improvement, including:

- **1) Agility**: respond with agility to new requirements for the separation of concepts and bringing the processes of IT and business; **2) Cost reduction**: promoting the reuse of services is possible to decrease development costs and maintenance, increasing efficiency. The share of services in a logic of coordination between, for example, sales, marketing, distribution, etc., can reduce costs and time in operational management; **3) Return of investment**: not only in the business itself, but also the capacity to boost future approaches, namely the simplicity of new systems.

2.2 Cloud Service Integration

In an era where Cloud Services invaded the form of both personal and business operation, a structural issue will be think of a SOA architecture that can be made available in the cloud without the use of heavy transformation mechanisms. Even in a logic of Private Cloud, technological or business expansion trend, will certainly lead to a strategy of scalable information systems. Moreover, web services are already used on the national integration Cloud for tax purposes, for example.

The current practice of application integration through APIs requires high coding effort and maintenance. Thus, the development and SaaS interactions with Web Services and with service-oriented architecture principles (SOA) for the implementation of business process will surely be a facilitator for the problems of migration to an approach of a cloud integration.

2.3 REST - Representation State Transfer

REpresentational State Transfer (REST) was originally presented as an architectural style for building large-scale distributed systems, whose principles have been used to explain the excellent scalability of the HTTP. According (Cesare, P. et al., 2008) The RESTful web services architectural style is based on four principles:

- **Resource identification through - URI**. A RESTful Web service exposes a set of resources, which identify the targets of the interaction with its clients. URIs identifies resources, which provide a global addressing space for resource and service discovery; **Uniform interface** - The resources are manipulated using a fixed set of four create, read, update, delete operations: PUT, GET, POST, and DELETE. PUT creates a new resource, which can be then deleted using DELETE. GET retrieves the current state of a resource in some representation. POST transfers a new state onto a resource; **Self-descriptive messages** – They must be decoupled from their representation, to be accessed in a variety of formats (e.g., HTML, XML, plain text, PDF, JPEG, etc.). Metadata about the resource is available and used, for example, to control caching, detect transmission errors, negotiate the appropriate representation format, and perform authentication or access control; **Stateful interac-**

tions through hyperlinks - Every interaction with a resource is stateless, i.e., request messages are self-contained. Stateful interactions are based on the concept of explicit state transfer. Several techniques exist to exchange state, e.g., URI rewriting, cookies, and hidden form fields. State can be embedded in response messages to point to valid future states of the interaction.

In the next figure (figure 1) its presented REST Web Service simplified architecture.

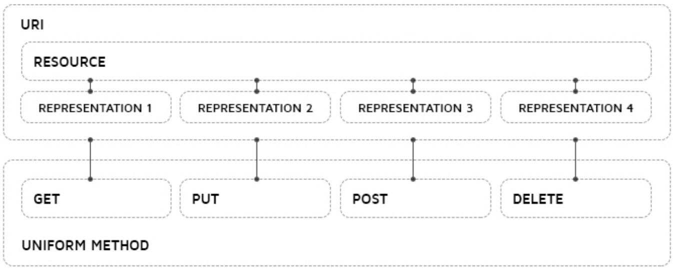


Fig. 1. REST Web Service simplified architecture

Also, there are six constrains that a RESTful architectural style requires must have: **Uniform interface** - servers and clients can interact, be modified and change without dependency as long as them interface remains the same; Client/**server loose coupling** - There is a clear separation of duty between client and server, i.e., does not matter to the client how storage is done, and does not matter to the server the client state; **Stateless**— There is no client state on the server. Client information is on the messages sent to server; **Cacheable**— Clients can store information in a temporary way; **Layering** — Servers do not know whether there are layers of abstraction between themselves and the end client; for example, whether they are passed through multiple security policies, APIs, and so forth; **Code on demand** — It should be possible to send custom functions to clients execute.

2.4 SOAP - Simple Object Access Protocol

SOAP is a standard protocol proposed by the W3C to interface Web Services, and that extends the remote procedure call (XML-RPC). Thus, SOAP can be considered as an evolution of XML-RPC protocol, much more complete and mature, that allows to perform remote procedure calls to distributed routines (services) based on an XML interface as interfacing language. (Castillo, P. et al., 2011).

On the technology level, SOAP is an XML language defining a message architecture and message formats, hence providing a rudimentary processing protocol. The SOAP document defines a top-level XML element called envelope, which contains a header and a body (Cesare, P. et al., 2008)) as presented on figure 2.

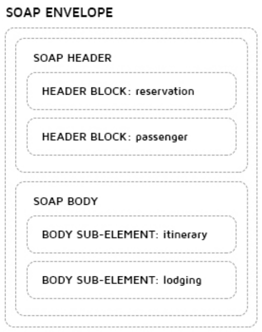

SOAP ENVELOPE
SOAP HEADER
HEADER BLOCK: reservation
HEADER BLOCK: passenger
SOAP BODY
BODY SUB-ELEMENT: itinerary
BODY SUB-ELEMENT: lodging

Fig. 2. Simple Object Access Protocol (SOAP) Envelop

Also according (Cesare, P. et al., 2008), the SOAP header is an extensible container for message a layer infrastructure information that can be used for routing purposes (e.g., addressing) and Quality of Service (QoS) configuration (e.g., transactions, security, reliability). The body contains the payload of the message. XML Schema is used to describe the structure of the SOAP message, so that SOAP engines at the two endpoints can marshall and unmarshall the message content and route it to the appropriate implementation. Web Services Description Language (WSDL) is an XML language for defining interfaces syntactically.

2.5 REST vs SOAP

A comparative analysis of SOAP / RESTful-based Web services of its key features is shown in Table 1.

Criteria	SOAP-based WS	RESTful-based WS
Server/ Client	Tightly coupled	Loosely coupled
URI	One URI representing end-point resource	URI for each the service
Transport Layer Support	All	Only HTTP
Caching	Not Supported	Supported
Interface	Uniform Interface (WSDL)	Non Uniform
Context aware	Client context aware of WS behavior	Implicit Web Service behavior
Data Types	Binary requires attachment parsing	Supports all data types directly
Method Information	Body Entity of HTTP	HTTP Method
Data Information	Body Entity of HTTP	HTTP URI
Describing Web Services	WSDL	WADL

Expandability	Not Expandable (No hyperlinks)	Expandable without creating new WS (using xlink)
Standards used	SOAP specific standards (WSDL, UDDI, WS-Security)	Web standards (URL, Http Methods, XML, MIME Types)
Security/Confidentiality	WS-security standard specification	HTTP Security

Table 1. Comparison of SOAP/ RESTful-based WS (AlShahwan, F. et al., 2010)

3 REAL LIFE SCENARIO

3.1 Context

As implementation scenario and proof of concept, we used a school information system, within the servers of Zendensino, Cooperativa de Ensino IPRL. It can be characterized as a school information system, with its beginning in 2000 and aggregator of all the pedagogical and management processes regarding vocational school in its main actors: teachers, students, guardians, class directors, administrative services, financial managers and pedagogical direction, providing in real-time, the full range of relevant information of the educational process, as shown in Figure 3:

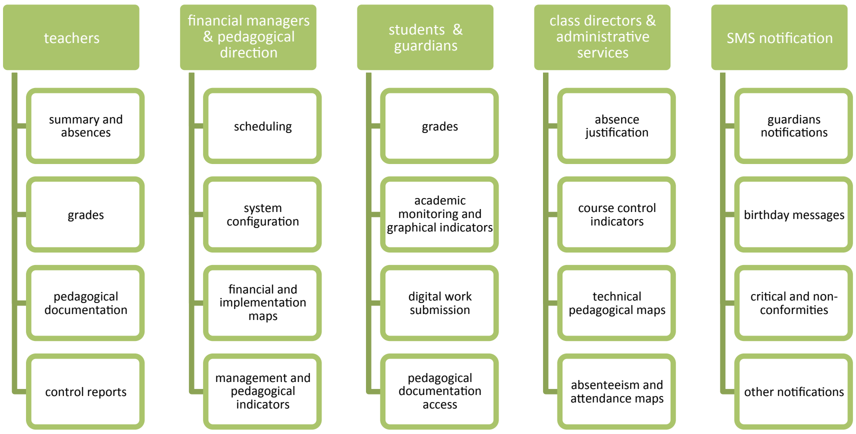

teachers
summary and absences
grades
pedagogical documentation
control reports
financial managers & pedagogical direction
scheduling
system configuration
financial and implementation maps
management and pedagogical indicators
students & guardians
grades
academic monitoring and graphical indicators
digital work submission
pedagogical documentation access
class directors & administrative services
absence justification
course control indicators
technical pedagogical maps
absenteeism and attendance maps
SMS notification
guardians notifications
birthday messages
critical and non-conformities
other notifications

Fig. 3. LPON (School Information System) modules

Over time, technology was implemented over a client-server architecture approach. At this time, contains different technologies (C, C # / PHP / Visual Basic, Sql Server, HTML, PHP, JQuery, Java Script and CSS). Creating another module e.g., for mobile environment, there is a need for further development of features that have been implemented for other modules and are not reusable.

It was chosen an SOA approach, exposing the necessaries resources through web services to be consumed, not only by the new module / prototype, but for all the components that can make use of it, now and in the future, over a web server (http/s).

3.2 Discussion

In this article, even taking into account some advantages of SOAP in relation to REST, the following characteristics were valued in order to choose RESTful among the possible approaches:

- **Simplicity** adaptation with a smaller learning curve. REST uses simple HTTP and standard commands - such as GETS, PUT, POST, and DELETE which leaves developers free to their own semantics and technical approaches; **Scalability**. With simpler component implementations and reduced complexity in the connection semantics, RESTful services can scale; **Familiarization** with the HTTP service (similarities with web technologies in designing approaches); The **acceptance of the https** security mechanisms as safe; There is **no need of expensive tools** to interact with the Web Service; **Performance** (no extensive processing required and the use of the existing web infrastructure) eliminates the need for an additional messaging layer, being efficient (REST with JSON can use less bandwidth).

3.3 Implemented MVC RESTful Server

The chosen framework to provide RESTful services was Laravel 4.0. Laravel is a PHP framework distributed under the MIT License. It follows the model-view-controller (MVC) architectural pattern, which enforces a separation between business logic from the input and presentation logic associated with a graphical user interface. The basic architecture is shown in the next figure (figure 4).

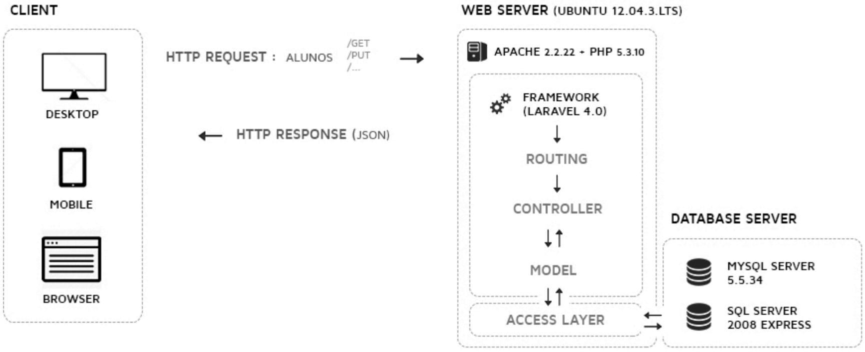


Fig. 4. Implemented MVC RESTful Server architecture

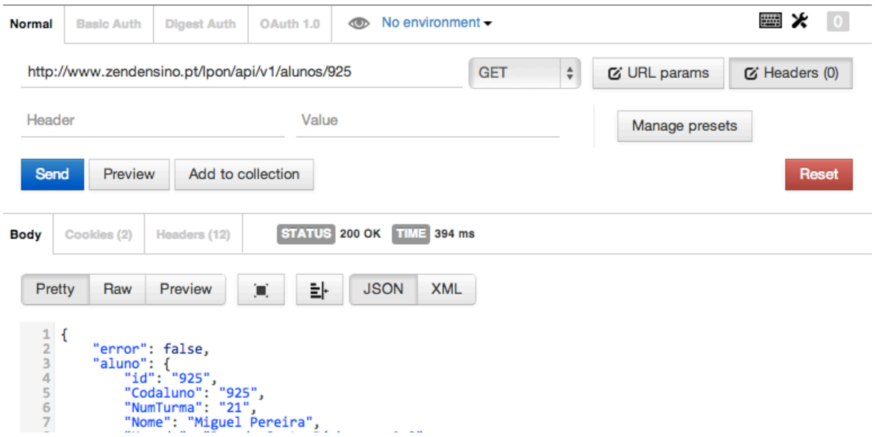
As an explanation, as a client sends a http/s request that is received by a web server. It s passed to the Laravel routing engine. Is then redirected to the appropriate controller class method, based on the routing URL pattern. The controller interacts with the model, which is a PHP object that represents an element of the application (such as a student, teacher, etc.) and is in charge of communicating with the database. After invoking the model, the controller then renders the final view (HTML, CSS and images) and returns the complete web page to the user's browser. In the presented implementation, as the goal was the released JSON, it was not used any viewer component. Instead, was developed an external interface application witch is responsible for consuming the returned JSON. Laravel promotes the concept that models, views, and controllers should be kept quite separate, by storing the code for each of these elements as separate files in separate directories.

3.3.1 Web Service Consuming – URI/GET, URI/PUT

Is now shown how the Web Service can be consumed (figures 5 e 6), using the browser Google Chrome and the POSTMAN extension as described on table 2.

<table>
<tr><th>Figure 5</th><th>Figure 6</th></tr>
<tr><td colspan="2">Resource: alunos</td></tr>
<tr><td colspan="2">URI: http://www.zendensino.pt/lpon/api/v1/alunos/</td></tr>
<tr><td>Operation: GET</td><td>Operation: PUT</td></tr>
<tr><td>Parameter: CodAluno</td><td>Parameter: nome</td></tr>
<tr><td>Function: List Students Data</td><td>Function: Update Student Name</td></tr>
<tr><td colspan="2">File: /app/controllers/AlunosController.php</td></tr>
<tr><td colspan="2">Authentication: Yes</td></tr>
</table>

Table 2. Legends of figures 5 e 6

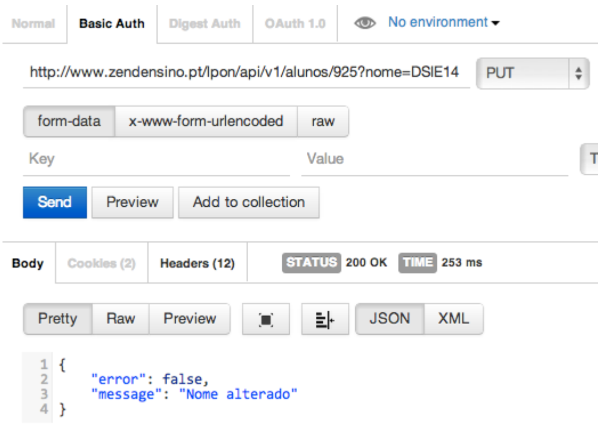


http://www.zendensino.pt/lpon/api/v1/alunos/925

STATUS 200 OK TIME 394 ms

```
{
    "error": false,
    "aluno": {
        "id": "925",
        "Codaluno": "925",
        "NumTurma": "21",
        "Nome": "Miguel Pereira",
```

Fig. 5. Consuming a Web Service sample – RESTful / GET



http://www.zendensino.pt/lpon/api/v1/alunos/925?nome=DSIE14 PUT

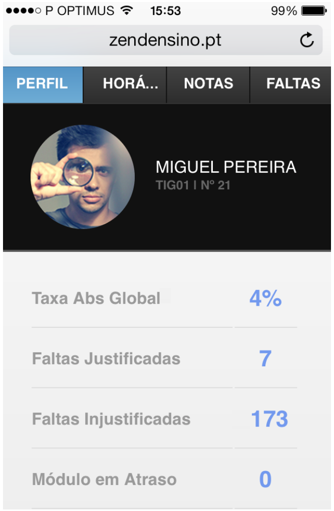
STATUS 200 OK TIME 253 ms

```
{
    "error": false,
    "message": "Nome alterado"
}
```

Fig. 6. Consuming a Web Service sample – RESTful / PUT

3.4 Client Example

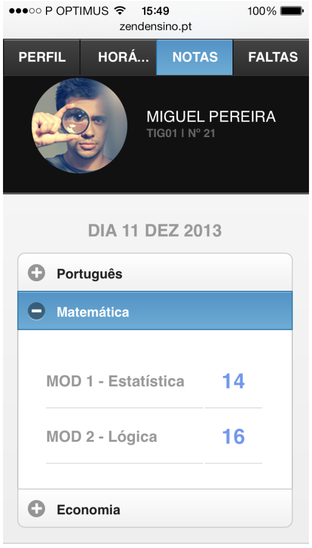
In the next 2 figures, as evidence of simplicity, its presented a simple mobile prototype client, developed in HTML, CSS and JavaScript in order to consume, after authentication (basic authentication), the representation of the resource Alunos state, regarding his overall situation (Figure 7) and his grades (Figure 8). This type of application can easily be distributed and adapted to mobile devices (e.g. IOS, Android, etc.) using e.g., PhoneGap, a free and open source framework that allows create mobile apps using standardized web APIs.



MIGUEL PEREIRA
TIG01 I Nº 21

Taxa Abs Global	4%
Faltas Justificadas	7
Faltas Injustificadas	173
Módulo em Atraso	0

Fig. 7. Student Global Status



MIGUEL PEREIRA
TIG01 I Nº 21

DIA 11 DEZ 2013

Português

Matemática

MOD 1 - Estatística	14
MOD 2 - Lógica	16

Economia

Fig. 8. Student Grades

4 CONCLUSIONS AND FUTURE WORK

As the future work, some issues that were not addressed should be studied, as: scalability, security, persistence and data access, endpoint resources, web application description language, expandability and data types, in order to get a closer view, regarding implementation and performance.

In this paper we have explored an SOA approach as a paradigm to take into account in the development of application in the context of a small business. In addition to the technical and business advantages, this approach can decrease the gap between business processes and technical in order to unify, under a logical service, business semantics.

Concepts and underlying technologies were explored in order to provide an integrated view of its components. Were also discussed implementation options and some suggestions were made, in order to prove that SOA approaches do not require complex models of development.

After a comparison between SOAP and REST approach, is suggested, in the context of small businesses, the use of REST. Finally, a demonstration prototype was presented, which included a Web server a PHP framework, the availability of state resources via Web Services through a RESTful MVC Server, and it was demonstrated, by example, how Web Services can easily be consumed and used by a simple Rest interface. It was also developed and presented a simple software application as GUI, with migration capability for mobile devices, and end point of the demonstration process.

In the presented prototype, it was possible to implement and demonstrate that the use of resources via Web Services in a SOA / Cloud approach, are not more heavily than traditional approaches, regarding time and technical effort, beyond the time that the learning curve entails.

As so, a SOA approach, with all of its advantages, is perfectly within reach of small businesses as add value. With the growth of cloud-based applications, and when we speak about Cloud, we are necessarily talking of services, a workable plan based on best practice, certainly will support the new business challenges, enhancing their technological sustainability.

5 References

1. AlShahwan, F. & Moessner, K. & Carrez, F., "Evaluation of Distributed SOAP and RESTful Mobile Web Services", International Journal on Advances in Networks and Services, 2010, vol 3 no 3 & 4

2. AlShahwan, F. & Moessner, K., “Providing SOAP Web Services and RESTful Web Services from Mobile Hosts”, Internet and Web Applications and Services (ICIW), 2010 Fifth, PP. 174 - 179.
3. Ashish, S., & Himanshu, A., & Ashim, R., "Integrating SOA and Cloud Computing for SME Business Objective", Transactions on Computer (WSEAS), 2012. PP 1–11.
4. Asif, M. & Majumdar, S. & Dragnea, R., “Hosting Web Services on Resource Constrained Devices”, ICWS 2007. IEEE International Conference on 9-13 July 2007. pp. 583-590.
5. Castillo, P. & Bernie, J. & Arenas M. & Merelo, J. & Garcia-Sanchez, P., "SOAP vs REST: Comparing a master-slave GA implementation",eprint arXiv:1105.4978,2011
6. Castro-Leon, E. & He, J. & Chang, M., "Scaling Down SOA to Small Businesses", IEEE International Conference on Service-Oriented Computing and Applications(SOCA'07),2007, PP 1–8
7. Cesare, P. & Olaf, Z. & Frank, L.,"RESTful Web Services vs. “Big” Web Services: Making the Right Architectural Decision", WWW 2008 / Refereed Track: Web Engineering - Web Service Deployment, 2008. PP 1-10
8. Delen, D. & Demirkan, H., “Decision Support Systems,” 2013, Decision Support Systems, vol. 55, no. 1, pp. 359–363, Apr.
9. Ivan, I. & Zamfiroiu, A., "Quality Analysis of Mobile Applications", 2011, Informatica Economica Journal, Vol. 15, No. 3, pp.136-152
10. Kim, Y. & Lee, K., “A Light-weight Framework for Hosting Web Services on Mobile Devices”, Web Services, 2007. ECOWS '07. Fifth European Conference on 26-28 Nov. 2007. PP. 255 - 263.
11. Lanthaler, M. & Granitzer, M., “SEMANTIC WEB SERVICES: STATE OF THE ART,” IADIS International Conference on Internet Technologies Society, 2010, pp. 1–8, Nov
12. Li, W. & Svärd, P., “REST-Based SOA Application in the Cloud: A Text Correction Service Case Study,” 2010, IEEE Congress on Services (SERVICES), pp. 84–90.
13. Muehlen, M. & Nickerson, J. & Swenson, K., “Developing web services choreography standards—the case of REST vs. SOAP,” 2005, Decision Support Systems, vol. 40, no. 1, pp. 9–29
14. Riva, C. & Laitkorpi, M., “Designing Web-Based Mobile Services with REST”, Service-Oriented Computing - ICSOC 2007 Workshops, Lecture Notes in Computer Science, 2009, Volume 4907/2009.

A Methodology Proposal for Reliability and Availability on WirelessHART: Best Practices

Mara Martins
FEUP, University of Porto
Porto, Portugal
{up201306980}@fe.up.pt

Abstract. Design of Wireless Sensor Network (WSN) need to address the failure of single or multiple network components. Most of these designs have stringent dependability requirements, where a system failure may result in economic losses, or damage for human life or the environment. In this paper we present a survey on approach to the design and validation of Wireless Sensor Networks aimed to optimize its reliability and availability requirements to WirelessHART when network devices are subject to permanent faults. From the results of the analysis we highlight that the WirelessHART Best Practices have different impact on the network.

Keywords: dependability evaluation; wireless sensor networks; fault tree analysis, wirelessHART.

1 Introduction

Commonly, in industrial environments, the applications are supported by wired communication networks. However, the industry has recently shown interest in moving part of the communication infrastructure from a wired to a wireless environment and are increasingly being used into critical application and consequently, makes the level of trust on WSN a factor of the utmost importance, directly linked to the success or failure of industrial large-scale WSN applications [1].

In this new paradigm, the main goal is to reduce various operating costs such as installation weight in the case of offshore platforms, scalability of the applications, maintenance, and also enable monitoring of new processes that are impractical with current communication technology. Another benefit is the flexibility of tests that can be performed more quickly reducing operating costs (operators, plant shutdown, keeping in risk areas). Taking this into consideration, WirelessHART indeed, becomes the best option to be applied as the communication solution for the last mile connection in process monitoring and control applications in industrial environments [5], [6], [1].

In communication networks, there are two types of failures that can be classified as either temporary or permanent. When a permanent fault occurs, they are caused by hardware failures, affecting the network devices and their repair is needed to make them available again. Transient faults normally impair communication links between network devices and are typically caused by noise or electromagnetic interferences. In the analyzed studies, only permanent faults was taken into consideration to have a greater impact on the functionality of the entire system. In such cases, instantaneously it detects the communication failure with the affected devices. In worst case scenarios, several network devices can become isolated, as a router failure, and as a consequence, it may take a system failure by various disturbances or deactivation of the control applications [5].

Important design decisions for evaluating the reliability of WirelessHART networks can be anticipated through the use of an appropriate Methodology. Namely, regarding its topology, criticality of the devices, redundancy aspects and network robustness [4]. This becomes very useful information specifically on the initial planning and design phases, and can be used during the system's life-cycle, e.g alternative paths to the gateway can be created to improve the overall reliability of the network.

In this paper we attempt to study and examine the dependability evaluation of the practices indicated by the HART Communication Foundation (HCF) presented in [4] and [5] when network devices are subject to permanent faults and analyze their performance.

The remainder of the paper is organized as follows: In Section 2 the aspects previous survey works are discussed. In Section 3 we describe briefly the WirelessHART networks when configured according to the best practices by the HART Communication Foundation (HCF). Section 4 describes the proposed methodology in [4] and [5] for the reliability and availability evaluation of Wireless Sensor Networks. Section 5 presents the analysis from the dependability evaluation, considering different fault scenarios and using the rules indicated by the HCF best practices. Finally, Section 6 concludes the paper.

2 Previous Work

The network reliability problem is a classical reliability analysis problem and it is known to be an NP-Hard. Constant research for a solution to the problem has been performed for wired networks [6], the author assumes failures of both hardware and software to measure the reliability and availability of a wired network, giving clear approach when there is occurrence of failures about the topology adaptation strategy and the state-space enumeration. The dynamism of the network is the major difference when comparing the reliability analysis of wired and wireless. Due to mobility of devices and higher frequency of failures of links, it can be stated that the dynamics of the wireless networks is greater. Studies by [A. Majdara and T. Wakabayashi, 2009], was one of the pioneers on the evaluation of reliability of a network for broadcasting. Was considered by the authors in this paper, only links and reliable devices and showed that the two-terminal reliability problem for radio broadcast networks is computationally difficult.

A proposed framework to the evaluate the reliability of a Wireless Sensor Network was described in [J. Song, S. Han, A. Mok, D. Chen, 2008]. The proposal was to develop a coverage-oriented reliability mechanism grounded on coverage requirements. If the total area of coverage attainable by a subset of nodes is greater than A, the network will fail, because there is no subset of nodes that are fully operational and whose own generated traffic can reach the sink [5]. The proposal above does not allow to characterize flexible failure conditions and also lacks support for add alternate devices and neither outline the critical level devices. However, it should be noted that this model has a better performance over the conventional 2-state (operate / fail).

In the proposed solution presented in [K. S. Trivedi and R. Sahner, 2009], the authors propounded a new mechanism using a different methodology for evaluating reliability of a Wireless Sensor Network through graphical representations in order to measure the reliability according to the amount of numbers of functional spanning trees. Thus, if there is at least one spanning tree, then, can be said that the network is reliable. This is not a complex solution and has a good applicability in the analysis of the topology control mechanisms. However it is not suitable to evaluate arbitrary WSN. In this solution, it is not possible to use and validate physical redundancy, neither to compute the criticality of the devices. Due the failure dependence for a spanning tree condition, the flexible failure conditions are also very hard to represent [5].

The proposals previously presented, provide only incomplete solutions to the network reliability problem. The approaches are restrictive to specific scenarios regarding the definition of the topology, industrial applicability, reliability metrics and other scenarios. Considering this detail, consistent studies on evaluation of WSN dependability are still performed, as the solution presented in [4] and [5].

3 WirelessHART

Following, is briefly presented the one of the main features of the standard WPAN network, the WirelessHART, specifically designed for use in industrial environments.

WirelessHART is the first open wireless communication standard approved by the International Electrotechnical Commission on September 2008 as a public specification available. It was designed for Wireless HART Communication

Foundation (HCF), created specifically for measurement and process control applications with the purpose of transmitting messages HART without the need to use the usual transmission channels.

The Standard WirelessHART foundation was based on the following main requirements: a) it must to be simple; b) scalable; c) reliable; d) secure; and e) support existing for HART technology [8]. In this Standard, eight types of devices are defined: network manager, network security, gateway, access point, field device, adapter, router and handheld device presented in Figure 1. All nodes connected to this wireless network (all devices) perform basic mechanisms to support network formation, maintenance, routing, security and reliability [4], [6].

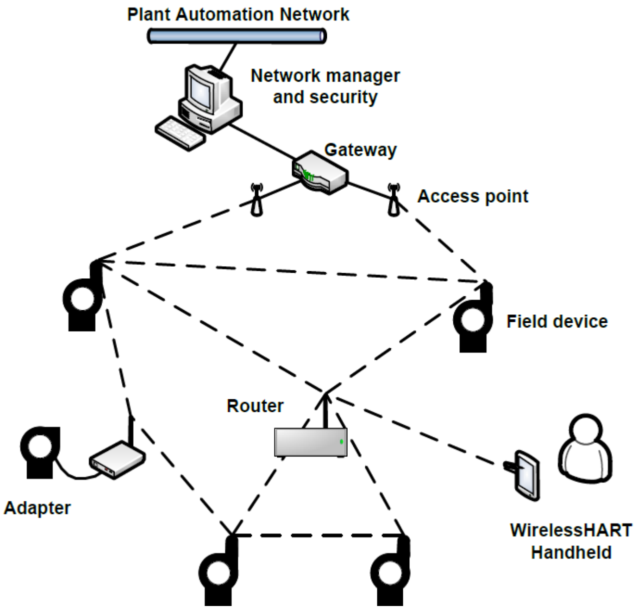


Fig. 1. WirelessHART devices.

The WirelessHART Architecture is based on the OSI 7-layer communication model, however, it is composed of five layers: physical layer, data link layer, network layer, transport layer and application layer. The physical layer is built on the IEEE 802.15.4 standard, enabling a WirelessHART radio belonging to the same baud rate and IEEE 802.15.4 radio modulations, but with a total of 15 channels used. The data link layer is responsible for providing many essential mechanisms for proper operation of the network and ensure reliable communication between jumps, maintenance data, QoS and scheduling. It is also in this layer where superframes and time dimension multiple access (TDMA) are used for access control settings. The network layer performs packet routing and addressing, besides providing secure communications between end-to-end links. The transport layer has the simple functionality to transmit reliable end-to-end data. Finally, the application layer, aims to fragment the data, and provide a high level interface for the users. In this layer features of HART technology are inherited, and all the actions and procedures are oriented commands. [5], [6].

4 Methodology for Reliability and Availability Evaluation

In this section, we provide a brief introduction to tree analysis concepts that are closely related with the proposed methodology and the dependability evaluation.

4.1 Fault Tree analysis Technique

Fault Tree Analysis (FTA) is a powerful technique for qualitative and quantitative evaluation of reliability and availability of systems that has emerged as an option for the previously discussed approaches. The FTA is a valuable tool that allows identification of possible accidents since it provides system failures, and can use it as a robust design tool. The success of this solution, is due to the acceptance of failure analysis of an undesired state of a system analyzed using Boolean logic (where the events occur or do not occur) to combine a series of lower-level events. [3], [7].

The most common FTA solution, generates automatically the fault tree according to the network specifications. But there are other FTA solutions that aggregate various modelings through representations of state transition tables, and also solutions to model a timed automaton or even generate the fault tree using digraphs. It is noteworthy that all the proposed FTA solutions apply dependency relationships between the system components for the generation of the fault tree. [2].

4.2 Dependability Evaluation of WirelessHART Best Practices

Usually, the WirelessHART applications has a smaller number of devices connected to the network, making it tractable to evaluate reliability of the network regarded as NP-hard. The proposed solution presented in [4] is discussed in this section and the methodology used is summarized in Figure 2.

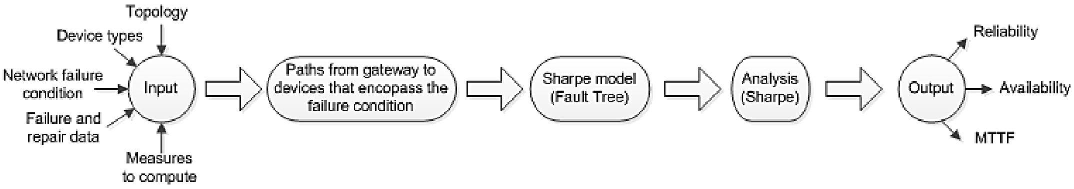


Fig. 2. Methodology for reliability and availability evaluation.

The proposed methodology for reliability and availability evaluation consists of 5 steps. First, the process starts with filling out information about the network topology, device types, network failure condition and device's failure and repair processes. The information referents to network failure conditions, define the conditions under which a network failure may occur. It is defined by a logical expression that sets the failure state of the field devices. The next step of the solution is to find all the paths of devices that compose the failure situation to be able to achieve the conditions of flexible failures and to support self-healing routing protocols. In step 3, a fault tree is generated using informations previously provided. At this stage, the metrics of interest calculations are performed using Sharpe tool. In the last step before the output, it is possible to assess the component's criticality factors, mean time to failure, device availability and reliability.

To automate the process of this approach was developed a specific tool to perform all the previous steps [4], [5]. The automation of this approach will not to be discussed in this case study.

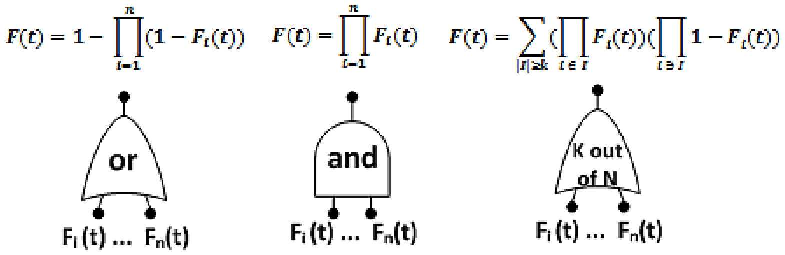

$F(t) = 1 - \prod_{i=1}^{n}(1 - F_i(t))$
$F(t) = \prod_{i=1}^{n} F_i(t)$
$F(t) = \sum_{|I|\geq k}(\prod_{i\in I} F_i(t))(\prod_{i\notin I} 1 - F_i(t))$
or
and
K out of N
$F_i(t) \dots F_n(t)$
$F_i(t) \dots F_n(t)$
$F_i(t) \dots F_n(t)$

Fig. 3. CDF for the gate output.

The FTA solution used in this methodology, composed by events and logical gates. Events may be performed under normal or faulty conditions, whether human-made faults, components failures or environmental conditions. These events are considered boolean. The logical gates such as AND, OR and K-out-of-N are used to represent the cause-effect relationship between events.

In Fault Trees (FT), TOP event indicates a system failure, from this event and proceeding backwards, the possible root causes are identified. Thus, this must be the first condition to be considered when building a FT. And when using probabilistic calculations, it is possible to calculate the probability of the TOP event by reference to the basic events probabilities. For each gate type, a different calculation is performed. Assuming a gate with *n* independent inputs (events), where the occurrence of event *i* is described by a cumulative distribution function (CDF) *Fi(t)*, as presented in Figure 3.

When an AND gate is used, a failure condition only occurs when all input events have occurred. However, when an OR gate is used, the failure condition only occurs if at least one input event has been occurred. Finally, if a gate K-out-of-N is used, the failure condition only occurs if at least k input events have occurred. When a FT does not contain any repeated event, the probability of the TOP event can be obtained through a direct calculation using the probabilistic formulas shown in Figure 3. If there are repeated events, the equations become invalid.

In the following sections, will be described briefly the necessary steps to create a model that represents a network failure.

4.1. Device failure condition

After creating the NFC, is essential to define conditions to leading to failure of the device field. For this purpose, two possibilities were considered:

i) Hardware failure: In case of hardware failures, it is assumed that an *i* device has a defect that prevents it from fully performing their duties. This situation is represented by Fd*i* event, shown in Figure 5, letter a;

ii) Conectivity problem (there is no valid paths): The connectivity failures are created when there are no paths between the device and the gateway. They are also represented by cp event, shown in the Figure 5, letter a.

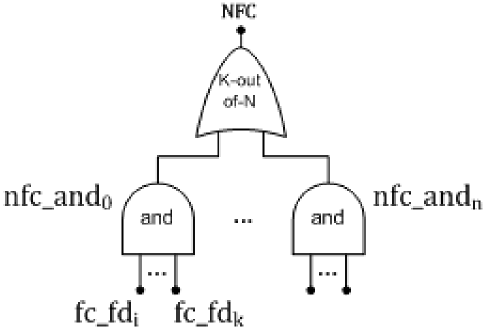


Fig. 4. Network failure condition.

To enable hardware failure events, it is necessary that all paths between the device and gateway have failed, as shown in Figure 5bFor a path be considered faulty, at least a device connection as gateway, access point, router, or field device must fails in some part of the route, as described in Figure 5c.

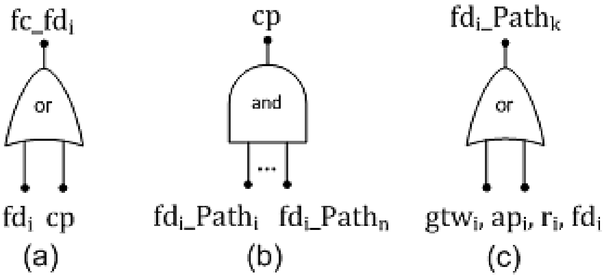


Fig. 5. Device failure condition.

4.2 SHARPE Model

The last step of the methodology created by [4], was transform the generated fault tree into a model suitable for SHARPE. This tool allows to compute several reliability and availability measures and obtain results from different aspects, either numerically or symbolically (using exponential polynomial terms) in particular instants of time (transient) or in steady-state [5]. Following, we briefly describe the input data necessary to compute these measures. Usually, after to get the fault tree, is necessary to replace all the basic events of the trees by availability or reliability functions [4].

5 Discussion

In this section the results obtained by [4] are detailed using the proposed methodology to evaluate the best practices suggested by the Hart Communication Foundation.

5.1. Scenario

The scenario chosen, consider a network Mesh WirelessHART in conformance with the rule of five and the rule of three. The topology chosen is shown in Figure 6. The fact that various topologies support the employability of the both rules, the author decided to use a symmetrical topology with the gateway at the center, placed according to a hexagonal pattern at fixed distances from the gateway. All straight lines connecting the devices represent communication paths [4], [5].

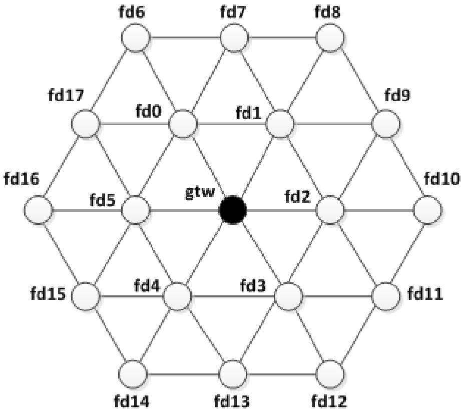


Fig. 6. Topology adopted.

The topology used contains 18 devices. Among these, 4 are fd9, fd11, fd15 and fd17, and consequently the network failure condition shall consider only the remaining 14 devices. Even in the scenario, it was assumed by the author that device failures occur with a constant rate and relative repair rates were only considered repairs, when evaluating availability, it assumes a constant repair.

5.1. Fundamental Rules

The HCF solution, provides three fundamental solutions to design WirelessHART Network.

5.12. Rule of Five

A WirelessHART network needs to have at least five field devices within the communication range of the gateway. This rule is evaluated by removing some connections with the gateway neighbors and measure the impact generated on the network [5]. Taking into account the topology of Figure 6, due to its symmetrical nature, any connection can be eliminated between the gateway and its neighbors to guarantee the rule of five as presented in Table 1.

Table 1. Rule Five Configurations

Num. neighbors (gateway)	Connections eliminated
6	-
5	$fd_0 - gtw$
4	$fd_0 - gtw$,$fd_1 - gtw$
3	$fd_0 - gtw$,$fd_1 - gtw$ $fd_2 - gtw$
2	$fd_0 - gtw$,$fd_1 - gtw$ $fd_2 - gtw$,$fd_3 - gtw$

To measure the impact of rule of five, the MTTF network computations are performed. Taking into account that MTTF (j) is the network MTTF When the gateway has j neighbors, then the impact of rule of five can be given by:

$$MTTF_{ratio} = \frac{MTTF(j) - MTTF(5)}{MTTF(5)}$$

The results of the measures of impact are shown in Figure 7. It is possible to see that the influence of rule of five has different impacts. When the number of devices which leads network is small, the influence of the number of neighbors is almost negligible and in other hand, it is possible to observe an opposite behavior. That is, the number of communication paths has a stronger influence. In these conditions adding neighbors to the gateway could be a useful solution to guarantee the dependability requirements of applications [4].

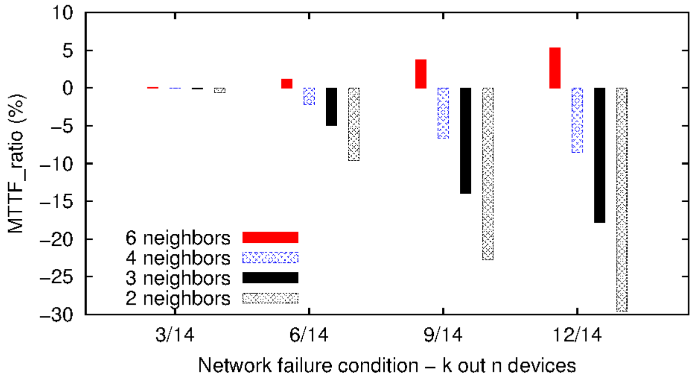


Fig. 7. Rule of five evaluation.

5.12. Rule of Three

The rule of three must have at least three neighbors. The evaluation considered the rule of five and choose a target device which is not a gateway neighbor. After that, we eliminate some connections of the target device with their neighbors and measure its impact on the network. Similarly to the rule of five, in the analysis of rule of three we use the network MTTT as the evaluation metric:

$$MTTF_{ratio} = \frac{MTTF(j)-MTTF(3)}{MTTF(3)}$$

The results of rule of three evaluation are presented in Figure 8. Differently from rule of five, it is possible to distinguish the influence of rule of three for all network failure conditions. It is possible to observe that the results follow a curve with a unimodal behavior.

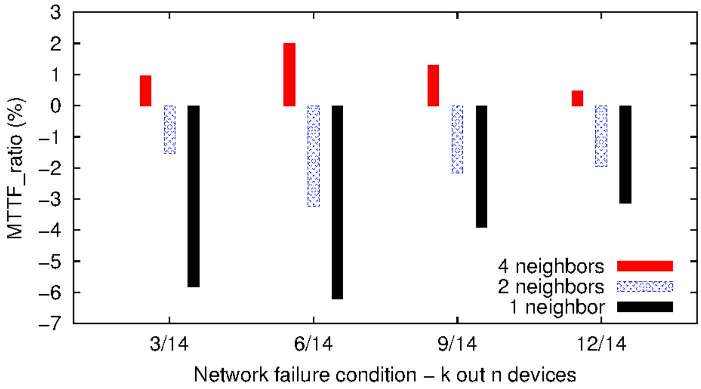


Fig. 8. Rule of three evaluation.

In general, the network MTTF decreases when the connections with the target device are eliminated. The reverse result is found when the connections are added. From this point of view, the rule of three and rule of five have the same behavior. For both rules, the impact to eliminate a neighbor is greater than add other neighbor.

Tests with the rule of 25% (when the network grows at least 25%) were performed and it was concluded that as expected, the MTTF network increases when the number of neighbors increases passage. However, increasing the number of neighbors gateway in real scenarios might be impractical due to the limitations of the environment.

6 Conclusions

The purpose of this study was to establish a state of the art on the subject of Wireless Sensor Networks, and more specifically in the study of a solution based on WirelessHART Best Practices. Through some solutions presented in Chapter 2, it is possible to realize that most of them provides incomplete solutions to the network reliability problem. Thus, the study in [4] used a previously proposed methodology to evaluate the dependability of Wireless Sensor Networks based on the rules indicated by the HFC.

Initially, it was possible to identify from the results that the best practices have different impact on the network. After removing and adding network connections in the three rules and the rules five, was possible to evaluate the real impact and assert that the impact to eliminate a neighbor is greater than add other neighbor for the both rules, and as the network grows the rule of 25% becomes dominant.

The advantage of this methodology is that the solution can be used as an analysis tool for network design and also to evaluate different best practices.

Acknowledgments. This project was financially supported by the CAPES, Brazil.

References

[1] Alessandro Testa.: Dependability assessment of wireless sensor networks with formal methods. Ph.D. Thesis, University of Napoles, Via Claudio 21, Napoli, Italy, 2013.

[2] Hou, W.: Integrated Reliability and Availability Analysis of Networks with Software Failures and Hardware Failures. Ph.D. Thesis, University of South Florida, Fowler Avenue Tampa, FL, USA, 2003.

[3] Ivanovitch Silva.: Uma Metodologia para Modelagem e Avaliação da Dependabilidade de Redes Industriais Sem Fio. PhD thesis, Faculty of Engineering, University of Porto. Natal, Brasil, January (2013).

[4] Ivanovitch Silva, Luiz Affonso Guedes, Paulo Portugal and Francisco Vasques.: Dependability Evaluation of WirelessHART Best Practices. In Proceedings of the 17th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA-2012), Krakow, Poland, September (2012).

[5] Ivanovitch Silva, Luiz Affonso Guedes, Paulo Portugal, Francisco Vasques.: Reliability and Availability Evaluation of Wireless Sensor Networks for Industrial Applications, Vol. 12 Issue 1. Sensors. (2012) 806-838.

[6] Jianping Song, Song Han, Aloysius K. Mok, Deji Chen, Mike Lucas, Mark Nixon.: WirelessHART: Applying Wireless Technology in Real-Time Industrial Process Control. In Proceedings of Real-Time and Embedded Technology and Applications Symposium (RTAS), Missouri, USA, April (2008).

[7] Latha Venkatesan, S. Shanmugavel, Chandrasekaran Subramaniam.: A Survey on Modeling and Enhancing Reliability of Wireless Sensor Network. Vol. 5 Issue 3. Wireless Sensor Network (2013) 41-51 11p.

[8] Mohammad Sadegh Javadi1 , Azim Nobakht2 , Ali Meskarbashee.: Fault Tree Analysis Approach in Reliability Assessment of Power System. VOL. 2, NO. 6. International Journal of Multidisciplinary Sciences and Engineering .September (2011).

[9] Tomas Lennvall, Stefan Svensson, Fredrik Hekland.: A Comparison of WirelessHART and ZigBee for Industrial Applications. In Proceedings of Real-Time and Embedded Technology and Applications Symposium (RTAS), Missouri, USA, April (2008).

Improving wireless sensor network efficiency in dense cluster

Benedito Bitencort

Faculty of Engineering, University of Porto – Porto - Portugal
benedito.junior@fe.up.pt

Abstract. In several monitoring applications are using clusters topology with autonomic behavior such as those can be applied in environment monitoring. It is widely used to improve the transmission over the network in applications when the sensor nodes must be thrown in the environment. A highly dense cluster has to be autonomic to treat lots of collisions; interferences caused by neighbors sensors; delayed messages and avoid missed deadlines to improve the network efficiency. This work presents a comparison between no autonomic probabilistic approach and a autonomic approach using genetic algorithm to improve the network efficiency into a high density cluster and considering time constraints required by the environment. Simulations results have shown an increase up to 5.4% on average the efficiency metric using genetic algorithms with ZigBee (IEEE 802.15.4).

1 Introduction

Hundreds or thousands nodes may be deployed into the area that is desired to monitoring and control. In general, sensors, communication wireless device, unit processing and small battery form each sensor node. [3]

Resource constraint as software and hardware is one of largest challenges of wireless sensor networks. Each sensor node have considerable limited energy, limited memory capacity and processing [1,2,3]. Some applications of these networks have high density, sensors nodes might fail and the network topology may change at any moment.

Wireless sensor networks are becoming object of studies by researchers around the world to apply it in factory floor, environment monitoring, military solutions and others. These kinds of applications have some particularities such as hazardous environment and time constraint [1].

Several applications requires hard or firm deadlines to validate the messages sent over network. The hard deadlines are more difficult to be employed due to interferences on the wireless communication caused by other devices or even others sensors nodes nearby. The most common kind of deadline is soft or firm. These kinds of deadlines are used in applications wherein missed deadlines does not causes to a catastrophe unlike hard deadlines applications.

The network can be composed by homogeneous node or nodes with different capacities and characteristics [2-4] and sensors can continuous coming and going at any time causing a highly dynamic network topology. Another important issue is the energy because the sensors use batteries to work and is necessary the optimization of power consumption.

Sundry approaches especially those for environmental monitoring use a cluster topology formed dynamically when the nodes are throw on the area. WSN has to be autonomic because nodes may fail or others nodes may enter into the cluster and is necessary to keep the network QoS.

Inside a cluster there are lots of problems when the sensors node attempts to send messages through the network: (i) neighbors may cause interference because it might want to send a message in the exactly moment that other are sending it message, (ii) in each interference a backoff (technique used by IEEE 802.15.4) is computed delaying the message, (iii) messages misses deadlines because round drift.

In this paper will be analyzed the behavior of wireless sensor network within a single, highly dense cluster with autonomic features using genetic algorithm to improve network efficiency. One of the assumptions in this paper is: if the efficiency is improved then it is possible to conclude that the problems previously defined have been minimized.

The efficiency metric is considered here as the mean of messages that arrives in cluster-head during a period of time defined by the cluster-head. Each period is called round and a group of rounds is called section time.

In a round the sensor nodes gather the environment information and send it to the cluster-head. After each section time the cluster-head compute a new probability to the next section based on previous efficiency metric. The new efficiency will be used as the probability for the next section time.

The probability computed by the cluster-head and sent to the sensor nodes is the threshold for each sensor decide if will gather the environment information or just sleep until the next round.

Briefly the main proposal by this paper is maximize the mean of messages that arrive up to cluster-head within a section time using genetic algorithm approach [7] to compute the probability in each section. The moment needed to compute the new probability is called here by checkpoint.

After simulation results the genetic algorithm improved the network efficiency up to 5.4% in all tested density comparing with the probabilistic approach. The difference between them is just the autonomic behavior.

This paper shows the genetic algorithms behavior on application layer comparing with a no autonomic probabilistic approach and it is organized as follow: First is presented this introduction with a briefly discussion about wireless sensor networks, autonomic computing and communication. In second section is presented the concepts related with Autonomic Computing and in the third section is presented the related works. The proposed model used in simulations with an explanation about the metrics and network's characteristics assumed here is presented in 4. After the simulations results was shown in section 5 comparing between two approaches probabilistic and genetic algorithm. The probabilistic approach is not an autonomic approach but is helpful to analyze the network behavior when all sensor nodes are using the network. Finally, into section 6 is shown the conclusions about this work and future works.

2. Autonomic Computing

It refers to the self-managing characteristics of distributed computing resources, adapting to unpredictable changes. Self-management systems need little or nothing human intervention. Therefore, this paradigm could be well applied into Wireless Sensor Networks with intention of improve the behavior of the sensors node in a collaborative network when it is used in a hazardous environment and human interventions are impossible. Then autonomic wireless sensor networks (A-WSN) can be helpful when the network need to self-management its computational resources.

Autonomic communication is other paradigms that may be applied into Wireless Sensor Network. Autonomic computing and autonomic communication shares the same objectives and principles. In this work, there are not any distinction between them. Autonomic communication and Autonomic computing will be treat as the same.

In this paper there were addressed only two characteristics of autonomic system as self-configuring and self-optimization.

Self-configuring: Autonomic systems should be self-configured and self-reconfigured autonomically according policies of higher level previously defined. When a new sensor node is inserted, the systems does not known when happened and the whole system has to adapt [11].

Into wireless sensor networks, this characteristic is very important because the dynamic behavior of the network and nodes forces the adaptability every time changing the network topology and the nodes' behavior.

Self-optimization: According [10] self-optimization is the most important characteristic of WSN, because it might insure network's survival for a long time. This kind of network must identify and analyze opportunities to improve its working; A-WSN may learn to make choices and synchronizations into the network's parameters [11].

3 Related Works

The proposed model has been based in previous research works and it has adopted star topology. This topology is part of ZigBee that uses IEEE 802.15.4 MAC and physical standard.

ZigBee is a standard applied at building Automation, remote controls, smart energy, health care, network devices and others. It was proposed by ZigBee Alliance for reliable, cost-effective, low-power wireless networking. ZigBee technology will probably be embedded in a wide range of products and applications across consumer, commercial, industrial and government markets worldwide. It builds upon the IEEE 802.15.4 standard[4] which defines the physical and MAC layers for low cost, low rate personal are networks. It also defines the network layer specifications for star, tree and peer-to-peer network topologies and provides a framework for application programming in the application layer. The following subsections give more details on the IEEE and ZigBee standards.

Based on ZigBee standard a lot of approaches has used star topology [2,4,5] inside clusters where sensors nodes reach the cluster-head with just one hop. This approach was used to minimize the energy consumption and interferences caused by neighbors sensors because inside a cluster can be applied algorithms to control the sensor nodes.

Round concept was shown in [5] and is helpful to discretize the monitoring time by defining the period of time which allow the control of the network sensors . It main goal of this concept is to maximize Wireless Sensor Network coverage area, where the monitoring phase is divided in equal rounds duration and here is called by section time.

In [9] has used cluster-tree topology in factory floor with time constraint. It was applied to avoid blind spots and communication faults caused by interference during the monitoring section. The author believes that the clusters formed not suffer interference from other clusters.

The work in [7] compare two approaches of genetic algorithms to maximize the network efficiency to attend the quality of fusion and deadlines constraints. The genetic algorithm was applied in the application layer and tries to find out the better threshold computed by the efficiency to keep the network active for a long time.

4 Model

The proposed model assumes a cluster with dozens or hundreds of sensor nodes sending their data periodically to the cluster-head. The cluster-head performs a data fusion, and send it for base station outside the cluster through others cluster-heads or directly. The sensor nodes is called source sensors here.

All source nodes reach the cluster-head just with one hop and all network nodes were threw randomly on the environment. In this work will be assumed that the cluster was formed using the algorithm proposed in [6], ensuring self-assembled cluster. After the cluster formation all source nodes gather the environment information and sends out the information to the cluster-head.

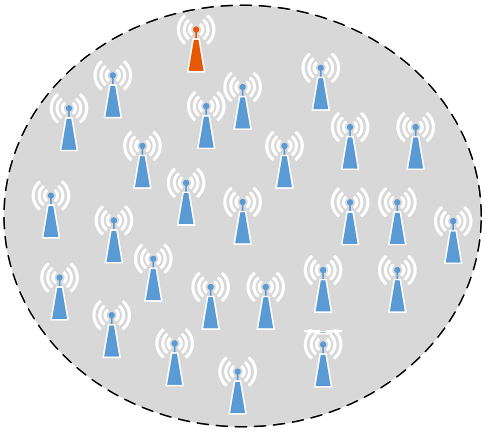


Fig. 1. Architecture adopted into the cluster. Each node reaches the cluster-head with just one hop.

Figure 1 is showing the network model adopted here and Table 1 summarizes the model characteristics assumptions. Informations are classified based on setup, communication, processing, connectivity and network addressing proposed by [7] and the protocol used in the simulations is the ZigBee, IEEE 802.15.4.

Table 1. Model characteristics [7]

Characteristic	
Setup	Composition: Homogeneous
	Organization: flat
	Mobility: Stationary
	Density: Dense
	Distribution: irregular
	Collection: Periodic with real-time constraint
	Propagation: Scheduled
	Connection: Simetric
	Channel allocation: dinamic
	Flow: broadcasting
Processing	Correlation
Connectivity	Single-hop
Addressing	Not applied

The goal of this work is not the data fusion algorithms, and then it assumes that this approach uses a simple mean to simulate a data fusion on cluster-head node as shown below in the equation 1. The total number of samples received and a(i) is an i-th sample received.

$$f(x) = \frac{\sum_{i=1}^{n} a(i)}{N} \qquad \textbf{(1)}$$

Sources nodes may be considered as low cost sensors, they can fail, and they gather homogeneous signal. Therefore, one of the few assumptions regarding the assumed data fusion algorithm is that this provides a better result when the number of samples used increases.

Considering the signal monitored it change every time and the data sent by the sources nodes has validity. In this paper will be considered firm deadlines to model the aging of monitored data.

4.1 Round, Sections and Checkpoint concepts.

The concepts used were Round, Section time and checkpoint and each one will be explained below.

Round: Each source nodes sends its messages periodically to each 1 second for cluster-head. This is done within the Round period. Round is a concept adopted by this model to mark the period for the source nodes gather the information about environment and send it to the cluster-head within deadline. The Round concepts is composed by the time to the sensor node wakes up, gather the environment information, send messages and sleep again.

Section Time: It is a group of rounds. Each section time is composed by 5 rounds and at the end of section time there is a checkpoint. The section time is the time on which the network is monitored. The figure 2 shown how it is divided.

Checkpoint: At the end of each section, a new probability (PE) is computed by the equation 2, and it is sent out for the source nodes to synchronize the new probability threshold, together with counter of round with the next round. The time for the checkpoint adopted here is equal to a one round.

$$PE = \frac{\max(ef)}{\lfloor f(x) \rfloor} \quad (2)$$

The equation 2 uses the maximum efficiency computed in each round within a section time by the estimation of source nodes in the network, computed by the equation 1. The equation 1 is used by equation 2 as the estimation of the number of source nodes active in the cluster. A floor is applied in the equation to round off to the previous value.

Another important assumption here is that the cluster-head does not known the number of source nodes inside the cluster. This characteristic is important because the autonomic algorithm has to be able to self-adjust according to the density fluctuation.

The cluster-head has a deadline to receive the messages equal to a 1 second minus ε value. This ε value is the time required to perform the data fusion and sleep until the next round. Therefore, for a given message be accepted by the cluster-head, the message has to carry the counter of round within current round. If a message does not arrive in time for the fusion process, this is counted as a missed deadline, and the message will not be dropped by the cluster-head.

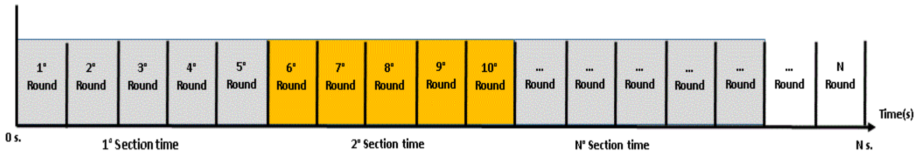


Fig. 2. A section is equal a five rounds. At the end of the each section, a new probability is computed.

The cluster-head and sources nodes are synchronized before the start each section within checkpoint time, and it is considered that the drift of the nodes' clocks are small enough to be no need for synchronization until the next session. But the messages can be out of date because the backoff algorithm implemented by ZigBee protocol.

A highly dense cluster has a lot of collision because the sources nodes try to send its messages and the medium are occupied. The probability sent by the cluster-head is helpful to minimize the quantity of source nodes that are sending messages in each round.

Moreover in a dense cluster, there is need to adjust the number of messages sent by the source nodes in order to establish a compromise between energy saving and improve the efficiency of the network to increase incoming messages by the cluster-head. This will be helpful to improve the number of received messages. For this reason, the proposed model assumes only one metric to be evaluated in the experiments: efficiency of the network.

A good question until here is about the setup's phase. In this phase, the cluster-head need to know the network and the only tool is the estimation about the number of sensor nodes deployed in the cluster than a broadcast message is sent out for all defining the PE equal to 100% and the counter of round equal 1.

4.2 The Efficiency (ef) metric

This metric measures the network behavior and its performance. If the efficiency is low, the number of collisions might have been high or the PE has not been well measured. Unlike if the efficiency is very high the PE has been well measured or many source nodes has drained its whole energy.

$$ef = \frac{\sum_{i=1}^{N} Mr(i)}{\sum_{j=1}^{m} Me(j)} \qquad (3)$$

The equation 3 shown how to efficiency (ef) is computed in this work. Mr(i) is the message received by the cluster-head within deadline into round "i" and Me(j) is the total number of messages sent during each round "j" that arrived up to cluster-head and within deadline or not.

In this paper is adopted two kinds of efficiency, the nominal efficiency and the real efficiency named by effective efficiency.

Nominal efficiency: It is computed using the real number of messages sent and not estimated.

Effective efficiency: will be better explained in the next section but the E. Efficiency is that computed by genetic algorithm using estimations.

5 Experimental Results

The simulation results show the behavior of the network accordance with the variation of density and two approaches are compared here; the probabilistic approach [8] is the first then second is an adaptive approach using genetic algorithm.

In Probabilistic approach each sensor node during time synchronization receives the PE computed and sent out by the cluster-head, then with the threshold defined the source node is able to decide if it can gather the information of environment and send it or merely sleep. In each period of time the source node computes a new likelihood and compares with threshold; if the probability generated by the source node is greater than threshold than it will sleep, in another case if it is less or equal the threshold then the source node sends out its information about the environment monitored.

The probabilistic approach has its PE computed by a random function with normal distribution ranging from 0 to 1.

In adaptive approach [7], the objective of the genetic algorithm was only increase network's efficiency. It computes the desired efficiency and sets the new transmission probability autonomically.

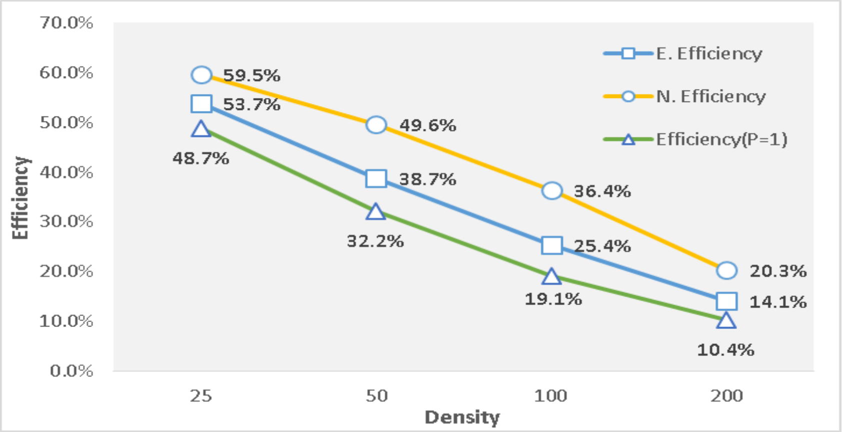


Fig. 3. Chart comparing the efficiency values obtained by simulation to the genetic algorithm.

When the density increases the efficiency decreases in both approaches, but slowly; However the network efficiency is better in adaptive approach. Effective Efficiency is the achieved real efficiency by the adaptive algoritms. In the figure 3 is shown the efficiency using the probabilistic approach with probability equal 100%, it forces all source node send its messages every period doing standard backoff proposed by IEEE 802.15.4.

The efficiency(P=1) forces all the sensor nodes send out the messages at a same time and a lot of collisions happen in the cluster and the ZigBee tryed to solve this problem with backoff algorithm. These values are important because we can use it as lower bound to the probabilistic and genetic algorithm appraches and the Nominal Efficiency can be seen as the upper bound.

The Nominal Efficiency is computed by estimation based on the number of messagens that arived in the previous section. The difference between Nominal Efficiency and Efective Efficiency is caused by genetic algorithm configurations because the cluster-head node does not know the quantity of sources nodes into the cluster then it has to estimate the desired efficiency value to set the new probability of transmission to the next sections. The distance between the effective and nominal efficiency into the figure 3 can be seen as the error of genetic algorithm's estimation.

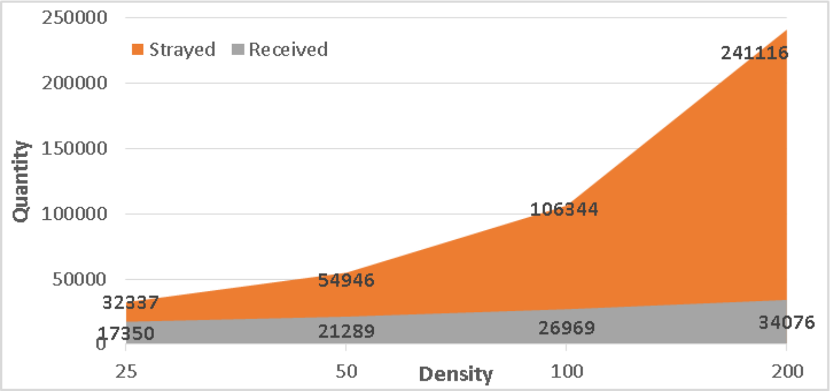


Fig. 4. Comparing Strayed and received messages.

A lot of messages arrive at cluster-head out of deadline. This is the reason because the falling efficiency. The missed and dropped messages are large comparing with the received and accepted messages. It is shown in fig.4.

Another important observation is about the round drift. If a source node has the round's counter out of date during a section time this sensor node can cause interference in the cluster by sending messages that will be dropped by the cluster-head.

The Table 2 accompanies the influence of probability variation by the genetic algorithm at some points of probability.

Until here is possible to realize that the efficiency of the network increases when the probability PE of transmission decreases, because interferences are decreased. The average column was computed using the quantity of messages arrived during a section.

Table 2. Influence of probability variation caused by genetic algorithm.

Probability (%)	Average	Efficiency (%)
100	16	32
32	10	62
20	7.2	73
14.4	5.5	77
11	4.4	81
8.8	3.6	85
7.3	3.1	86

Simulation results shows that the genetic algorithm was able to improve on average 5.4% network efficiency comparing between densities used here. An important behavior until here is about the genetic algorithms during the simulations, it found out a range of values for the probability and it varied within this range of values over the simulations.

6 Conclusions

This work shown a comparison between two approaches: a no autonomic approach and genetic algorithm approach. In probabilistic approach the probability parameter was handled without one objective and the algorithm does not learn anything about the network. But the second approach based on machine learning using genetic algorithm shown better results.

The assigned goal to the genetic algorithm was reached and the network efficiency was increased up to 5.4% on average. It is evident from this work that the approach proposed here showed relevant results.

Moreover by handling the probability with an autonomic approach was possible to minimize the (i) interference caused by the neighbors because merely a few sensors would send the messages; (ii) the message's delay caused by the ZigBee's backoff was also minimized and (iii) the messages does not missed its deadline improving the network efficiency.

The large amount of missed messages allows concluding that some sensors wasted a large amount of energy and until here is possible to realize that time constraints required by the application caused the rejecting by the cluster-head.

Therefore, in the future works will be evaluated how much energy is consumed by the genetic algorithm and if it may help to improve the quality of fusion (QoF) and the quality of service (QoS) into the Wireless Sensor Networks. A new goal for the genetic algorithm is try to find out the better probability to ensure the best level of data fusion at the same time ensures the best network efficiency.

Acknowledgments. The author thanks by the financial support from CAPES – Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior – Brazil.

References

1. Marrón, P., Minder, D., Embedded WiSeNts Research Roadmap, Information Society Technologies, IST-004400, November 2006.
2. Akyildiz, I. F., Su W., Sankarasubramaniam Y. e Cayirci E. (2002). A Survey on Sensor Networks. IEEE Communications Magazine, pp. 102-114.
3. IIyas, M. e Mahgoub, I. Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems, 2004, CRC Press Inc., Boca Raton, Florida USA.
4. Stankovic, J. A., Abdelzaher, T. F., Lu, C., Sha, L., Hou, J.C., Real-Time Communication and Coordination in Embedded Sensor Networks. Proceedings of the IEEE, Vol. 91, N°.7, (Jul.2003), 1002-1022.
5. T. Yan, T. He, J.A. Stankovic, Differentiated Surveillance for Sensor Networks, Proceedings of 1th International Conference on Embedded Networked Sensor Systems, 2003.
6. Sohrabi, K.; Merrill, W.; Elson, J.; Girod, L.; Newberg, F. & Kaiser, W. J. (2004), 'Methods for Scalable Self-Assembly of Ad Hoc Wireless Sensor Networks.', *IEEE Trans. Mob. Comput.* 3 (4) , 317-331 .
7. Junior, B. R. B. Fusão de dados Paralela em Rede de Sensores sem Fio Densas Usando Algoritmos genéticos. 2008. 97 f. Dissertação (Mestrado em Automação e Sistemas) –

Departamento de Controle, Automação e Sistemas, Universidade Federal de Santa Catarina, Santa Catarina, 2008

8. Pinto, A., Bitencort, B., Cabreira, U., Ribeiro, M., & Carlos, M. Fusão de Dados Tempo Real em Redes de Sensores Sem Fio Multimídia.*Webmedia,(Gramado: 2007). Anais do Simpósio Brasileiro de Sistemas Multimídia e Web*, *1*, 95-102.
9. M. Felske, C. Montez, A. Pinto, F. Vasques, P. Portugal, GLHOVE: A Framework for Uniform Coverage Monitoring using Cluster-Tree Wireless Sensor Networks, 18th IEEE International Conference on Emerging Technologies & Factory Automation, Cagliari, Italy, September, 10-13, 2013.
10. O'Hare, G.M.P., O'Grady, M.J., Marsh, D., Ruzzelli A. G., Tynan, R.; Autonomic Wireless Sensor Networks: Intelligent Ubiquitous Sensing. In proceeding of ANIPLA, International Congress on Methodologies for Emerging Technologies in Automation, 50 anniversary of ANIPLA, the Italian National Association for Automation, 2006.
11. Mainsah, E.; Autonomic Computing: the next era of computing. In: Electronics & Communication Engineering Journal, vol.14, n° 1, pp. 2-3, 2002.

Here is shown how the sources nodes works.

```
  While(SynchronizationMessagesNotArrive()){
      getSynchornizationMessage();
  }
  If(Monitor.getProbability() <= threshold){
      Monitor.setDataFromEnvironment();
      Packet.create(getDataFromEnvironment(),getAtualRo
und());
      Packet.sendPacket();
  }
  Sensor.SleepUntilNextRound();
  setAtualRound(getAtualRound() + 1);
```

This simple algorithm shown how the implementation in cluster-head was done.It is only to understang how it works.

```
  Start.setChromosomes();
  Start.setBudget(1000);
  Classifier.setCondition();
  Classifier.chooseAction();
  If(efficiency was improved){
      Banco.chooseChromosome();
      If(isCheckPoint){
            AG.chooseParents();
            AG.doCrossing();
            If(timeToMutation){
                  AG.doMutation();
            }than{
                  Classifier.newClassifier();
            }
      }than{
            Classifier.setCondition();
            Classifier.chooseAction();
      }
  }
```

SESSION 3

SOFTWARE ENGINEERING & SECURITY SYSTEMS

Model-based test case generation for Web Applications

Miguel Nabuco and Ana Paiva

Information Systems Security Audit – A real life scenario and prototype solution

Pedro Sousa

Support for automated teamwork submission leading to automated plagiarism detection

Rodolfo Matos

Model-based test case generation for Web Applications

Miguel Nabuco, Ana C. R. Paiva

Departamento de Engenharia Informática
Faculdade de Engenharia da Universidade do Porto
Porto, Portugal

Abstract. This paper presents a test case generation and filtering approach to test Web Applications using their models. The models are obtained from the users execution traces using a reverse engineering tool. Firstly, Test Paths are generated from the models. This is the number of possible different interactions a user could perform. Some filters are applied to the Test Paths to reduce the number of test cases. The Test Paths are then expanded into a tree structure, which will contain all the test configurations. A further filtering is applied which will provide the final test suite. This approach was evaluated with some experiments over three sample Web Applications.

1 Introduction

Web applications are getting more and more important. Due to the stability and security against losing data, the cloud and cloud-based web applications are progressively gaining a bigger userbase. Web application can now handle tasks that before could only be performed by desktop applications [20], like editing images or creating spreadsheet documents.

Despite the relevance that Web applications have in the community, they still suffer from a lack of standards and conventions [21], unlike desktop and mobile applications. This means that the same task can be implemented in many different ways. Pure HTML, Javascript, Java and Flash are just some of the technologies that can be used to implement a certain feature in a web application. This makes automatic testing for Web harder, because it inhibits the reuse of test code.

However, users like to have a sense of comfort and easiness, even some familiarity, when they use their applications. For this purpose, developers use some common elements, such as User Interface (UI) Patterns [23]. UI Patterns are recurring solutions that solve common design problems. When a user sees an instance of a UI Pattern, he can easily infer how the UI is supposed to be used. A good example is the Login Pattern. It is usually composed of two text boxes and a button to send the username and password data, and its function is to allow a user to access its private data.

Despite a common behavior, UI Patterns may have slightly different implementations along the Web. However, it is possible to define generic and reusable test strategies to test them after a configuration process to adapt the tests to those different possible applications. That is the main idea behind the Pattern-Based GUI Testing [22] (PBGT) project in which context this research work is developed. In the PBGT approach, the user builds a test model containing instantiations of the aforementioned test strategies (UI Test Patterns) for testing occurrences of UI Patterns on Web applications.

The goal of the work described in this paper is to improve the process of generating and filtering test cases to test the Web applications from their test models. The rest of the paper is structured as follows. Section 2 provides a state of the art of test generation techniques, addressing related works. Section 3 presents an overview of the PBGT approach, setting the context for this work. Section 4 presents the approach to generate and filter test cases. Section 5 presents some preliminary results and Section 6 takes some conclusions, summing up the positive points and limitations of this approach.

2 State of the Art

A good number of different test case generation techniques have been developed and researched in the past years. These techniques rely on different types of artifacts: some use software models, like Finite State Machines (FSM), others use information about the input data space, others use information from the software specifications. None of these techniques is exclusive. They can be combined to provide better results.

This paper focuses on model-based testing, which is a technique for generating a suite of test cases from models of software systems [1]. The tests are generated from these models so if they are not consistent with the software they represent, the tests will not be able to fully validate the application under test (AUT).

This testing methodology is very broad. Some approaches use axioms to build state based models using pre- and post- conditions [3]. Dick [2] developed a method to generate test cases from VDM models based on pre- and post- conditions. He created transitions models where each pre- and post- condition was represented as a state. If there was a variable x that lead to two states, *S1* and *S2*, when applied as argument in a function *f()*, a transition was created between the two states *S1* and *S2* ($S1 \rightarrow pre(f(x))$ and $post(f(x)) \rightarrow S2$). After this state machine was created, test selection techniques (such as random testing or state coverage) was allied to filter the number of tests to be performed on the AUT.

Using a slightly different approach, Claessen developed QuickCheck [4], a combinator library that generates test cases. In QuickCheck the programmer writes assertions about logical properties that a function should fulfill; these tests are specifically generated to test and attempt to falsify these assertions.

There are also some approaches that use FSMs. In FSM approaches, the model is formalized by a *Mealy machine*. A *Mealy machine* is a FSM whose output values are determined both by its current state and current inputs [5]. For each state and input, at most one transition is possible.

The main goal of FSM based testing tools is not complete coverage. To filter the tests to be performed, they use structural coverage criteria, such as transition coverage, state coverage, path coverage, amongst others.

Rayadurgam [8] presents a method that uses model checkers to generate test sequences that provide structural coverage of any software artifact that can be represented as a FSM.

Since most of the real software systems cannot be fully modelled with a simple FSM, most approaches use extended finite state machines (EFSM) [6]. A typical implementation on an EFSM is a state chart.

Frohlich [7] automatically creates test cases (which he defines as sequences of messages plus test data) from UML state charts. These state charts have originated from use cases which contain all relevant information, including pre and postconditions, variations and extensions. To generate the test cases, state charts elements are mapped to a planning language.

Similar to FSMs, labeled transition systems (LTS) are also being used to generate test cases. The difference between LTS and FSM is that LTS is non-deterministic: the same trace can lead to different states. Also, the number of state and transitions is not necessarily finite and it may not have a "start" or "end" states.

There is a particular testing theory for LTS systems, called IOCO (input/output conformance) [9]. The theory assumes the AUT to be an input enabled LTS which accepts every input in every state. There are some test generation tools that implement the IOCO theory, such as TVEDA [10], TGV [11], the Agedis Tool Set [12], and TestGen [13].

An alternative approach to IOCO is *alternating simluation* in the framework of *interface automata* (IA) [14]. IA does not require input completeness of the AUT, as opposite to IOCO. The IA approach was deeply refined and composes the foundation of Microsoft Spec Explorer Tool [15].

Some works were developed to provide test selection strategies on top of these frameworks. Feijs [16] implemented test selection based on metrics for IOCO and Nachmanson [17] implemented test selection based on graph traversal and coverage for IA.

However, for the testing of web applications, none of these approaches seemed adequate enough. A structure to allocate all the information containing multiple configurations for the same UI elements was needed [24]. Therefore, a new approach, built on a domain specific language, was created. The description of this approach can be seen in Section 3.

3 PBGT Overview

Pattern Based GUI Testing (PBGT) is a testing approach that aims to increase the level of abstraction of test models and to promote reuse. PBGT [18] has the following components (see also Figure 1):

- **PARADIGM** — A domain specific language (DSL) for building GUI test models based on UI patterns;

- **PARADIGM-RE** — A dynamic reverse engineering approach and tool to generate test models from existent Web applications.
- **PARADIGM-ME** — A modelling and testing environment to support the building of test models;
- **PARADIGM-TG** — A test case generation tool that builds test cases from PARADIGM models;
- **PARADIGM-TE** — A test case execution tool to execute the tests, analyze the coverage and create reports;

In Figure 1, the activities (rounded corner rectangles) with the human figure mean that they are not fully automatic requiring manual intervention. The activities with the cog mean that part (or all) of that activity is automatic. The numbers within the activities define their sequencing.

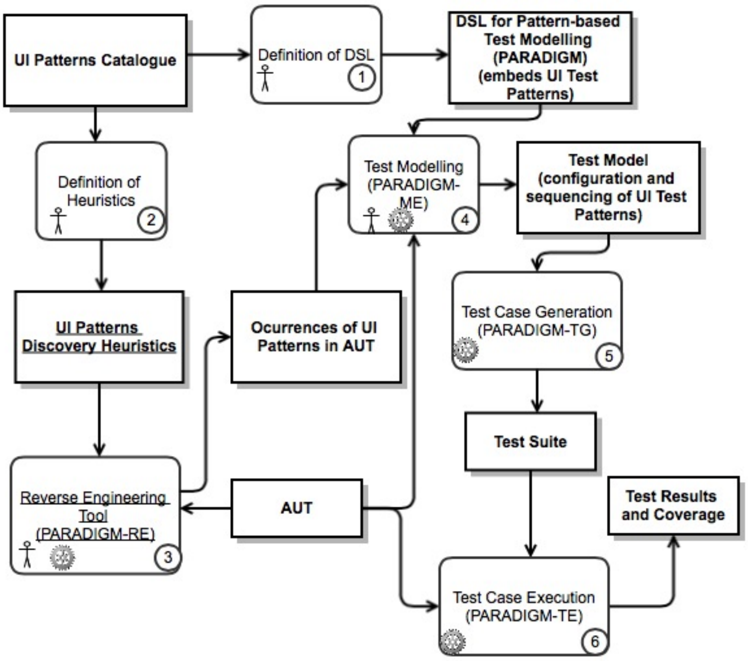

UI Patterns Catalogue
Definition of DSL
1
DSL for Pattern-based Test Modelling (PARADIGM) (embeds UI Test Patterns)
Definition of Heuristics
2
Test Modelling (PARADIGM-ME)
4
Test Model (configuration and sequencing of UI Test Patterns)
UI Patterns Discovery Heuristics
Ocurrences of UI Patterns in AUT
Test Case Generation (PARADIGM-TG)
5
Test Suite
Reverse Engineering Tool (PARADIGM-RE)
3
AUT
Test Results and Coverage
Test Case Execution (PARADIGM-TE)
6

Fig. 1. PBGT overview

Currently, PARADIGM is used to test web applications, but it can be easily expanded to test desktop and mobile applications. PARADIGM-RE starts by obtaining UI patterns [19], like login or search patterns, from the execution traces of the users navigating through the web application. RE not only

obtains the UI patterns, but also their configurations (for example, for a login pattern this includes the set of usernames and passwords used), resulting in **UI Test Patterns**.
This set of UI Test Patterns can then be exported to PARADIGM-ME, the modelling tool. In ME, the tester can build the whole web application model, by completing the UI Test Patterns information, defining the expected results and pre-conditions of each pattern. Then, the user defines the order in which they are performed by sequencing the UI Test Patterns with connectors. At the end of this step, the user should have a model of the web application as a whole, which elements can be performed in what order. In Figure 2, a model of Tudu (an online application for managing a todo list) is shown.

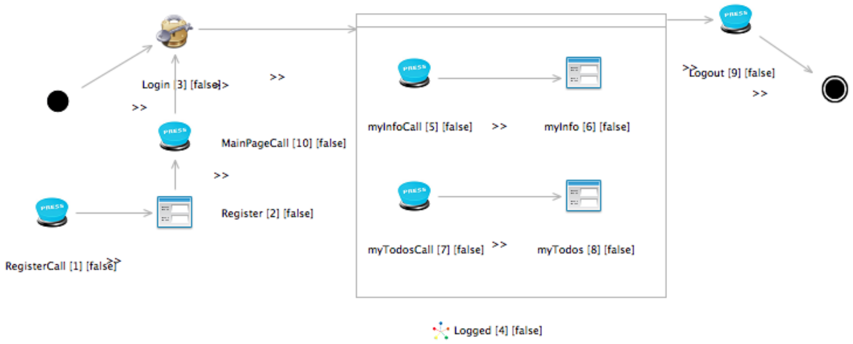


Fig. 2. Model of the web application Tudu

With the model fully completed, PARADIGM-TG will generate test cases from the model. The test case generation is the focus on this paper, and it fully described in Section 4. With the test cases generated, PARADIGM-TE will perform the tests on the web application and provide reports with the results.

4 Test Case Generation

In this section, it will be explained how the test cases are generated and filtered from the web application model previously generated and completed.

4.1 Path Generator

Each UI Test Pattern (which will be referred from this point onwards as "element") contains a specific ID, a number that identifies the element in the model.

A path is a sequence of elements that define a possible execution trace. A model can generate many paths, based on the connectors between elements. Also, some elements can be contained in groups and be performed in any possible order.
As an example, consider a web form where the user has to input his name (ID=1), age (ID=2) and e-mail (ID=3). These elements can be filled in any order. The number of paths will then be 3!, every possible combination between the three elements.The paths generated will be the following:

[1,2,3]
[1,3,2]
[2,1,3]
[2,3,1]
[3,1,2]
[3,2,1]

Each of these paths compose a Test Path, that can contains multiple test cases. So, the first step of the test case generation process is to generate the Test Paths from the model.
In case the model is too big or the testing goal is to test a single component and not the whole web application, a path filter component was implemented to reduce the number of Test Paths generated.

4.2 Path Filter

The path filter will restrict the number of test paths generated based on three simple conditions:

- **Cycles** — A Test Path can have cycles, i.e., a sequence of elements that can repeat itself multiple times. It is possible to reduce the number of Test Paths created by limiting the number of times each cycle is performed. In Figure 3, a cycle can be seen in the elements 1,2.
- **Mandatory and Exclusion Elements** — If the tester wants to test a certain feature or a certain workflow that can only be reached following a certain path, he can explicitly define the mandatory and exclusion elements. By excluding one element (or one sequence of elements), every Test Path that contains that element will not be generated. By excluding an element, other elements may be automatically excluded. In the model represented in Figure 3, if the element with the ID 4 is excluded, there will be no Test Path generated with elements 5 and 6.
- **Random** — The tester can also define a maximum number of Test Paths to be randomly generated. These Test Paths were already filtered by the two conditions described above, so this is the last condition to be applied.

4.3 Path Tree

After the initial Test Path Filtering, each Test Path must be decomposed in multiple test cases.

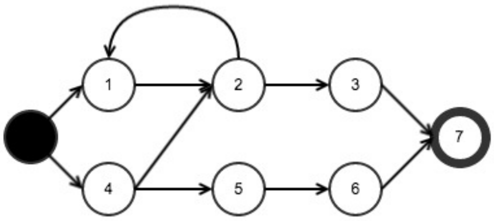


Fig. 3. Sample Test Model

One element can contain multiple configurations. A Login element can contain multiple valid and invalid configurations and each one of these can lead to different outcomes. Therefore, each configuration must generate a different test case. Considering a Test Path with x elements, each with y_x configurations, the total number of Test Cases generated will be:

$$Tc = y_x \times y_{x-1} \times ... \times y_1$$

To make the test case generation easier, a tree structure was created, as seen in Figure 4. Note that, if the element being inspected is a Login UI Pattern with an invalid configuration, an extra node will be created (Node 3), to create a test case that simply ends there (in the cases where valid credentials are needed to access private content, it is important to have this test case to check if the access is denied).
The tree is then travelled by height. An iterator starts from the tree root node and follows one branch until the end. Then, it marks the last node visited, saves the branch as a test case and starts again from the beginning, never travelling into the nodes already marked as visited.
From the Figure 4, the travelling order would be:

[1,4,1,5,2,6,2,7,3]

Creating the following test cases:

[1,4]
[1,5]
[2,6]
[2,7]
[3]

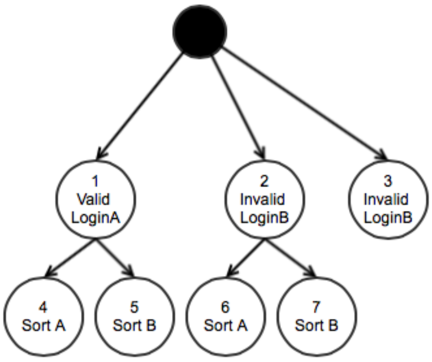


Fig. 4. Sample Test Tree

4.4 Test Case Filtering

From the example above, it can be seen that from two simple Login and Sort UI Test Patterns with two different configurations, 5 test cases are generated. If this is scaled to real-life applications, there can be a test case explosion, with too many test cases for the time and computational resources available. Therefore, just like a filter was applied to the Test Paths, a different filter has to be applied to reduce the number of test cases. The metrics applied in the filter were the following:

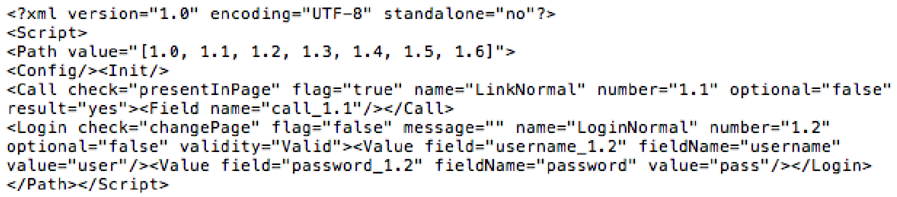
- **Specific configurations** — If the tree is filtered so that only test cases with specific configurations are considered, the number of test cases to be performed can drastically reduce. In Figure 4, if it is specified that only valid logins are of interest, the total number of test cases drops from 5 to 2, decreasing over 50%. The modelling tool provides enough flexibility so that the tester can generate test cases for the specific components he wants to test.
- **Random** — As with the Path Filtering explained in Section 4.2, the tester can also specify a number of test cases to be randomly chosen from the total set.

4.5 Exporting to PARADIGM-TE

After all the filters are applied, and the test cases are selected, the tool will then export the test cases into XML files, so that the test execution tool (PARADIGM-TE) can perform them and generate reports. Each Test Path (with several test cases) is stored in a different XML file.

This allows an easier test execution and debug, as it is easier to know if there is any Path not correctly defined.

An example of a XML file of a simple two nodes test case can be seen in Figure 5.



```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<Script>
<Path value="[1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6]">
<Config/><Init/>
<Call check="presentInPage" flag="true" name="LinkNormal" number="1.1" optional="false"
result="yes"><Field name="call_1.1"/></Call>
<Login check="changePage" flag="false" message="" name="LoginNormal" number="1.2"
optional="false" validity="Valid"><Value field="username_1.2" fieldName="username"
value="user"/><Value field="password_1.2" fieldName="password" value="pass"/></Login>
</Path></Script>
```

Fig. 5. Two nodes XML file

5 Evaluation

This approach was tested in three different web applications: **Tudu** (Figure 2), a web application for managing todo lists; **iAdressBook** (Figure 6), a PHP address book application; and **TaskFreak** (Figure 7), a task management web application. Each model contained several configurations for each element.

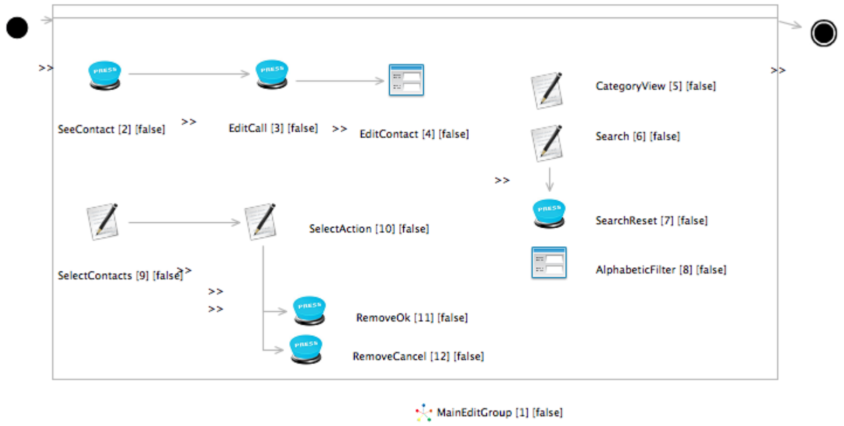


Fig. 6. Model of the web application iAddressBook

The Test Execution tool (PARADIGM-TE) is still not implemented with the test generation making it not fully possible to measure the impact of the test filtering.
However, from the experiences made, it was possible to see clear advantages of this approach:

- **Distributing Test Suites** — Since the test generation tool can generate specific configurations, the testing process can be divided so that each

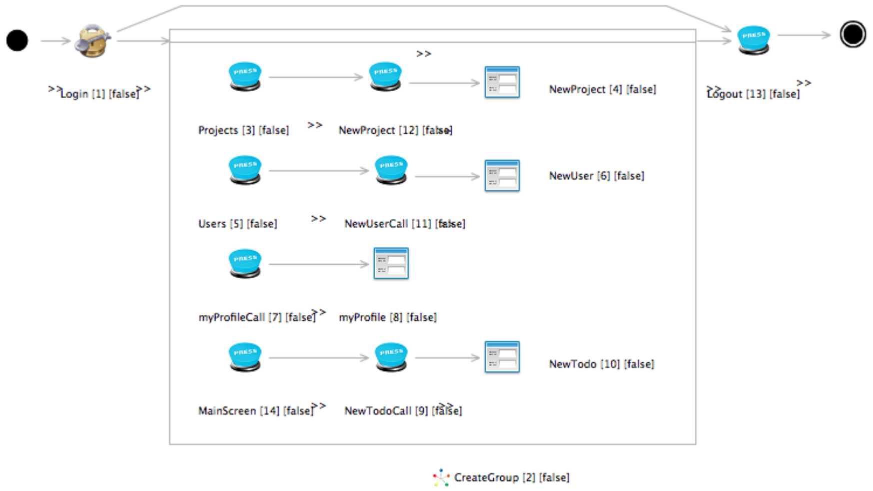


Fig. 7. Model of the web application TaskFreak

workstation runs a specific Test Suite. In this way, all the tests would still be performed on the application under test (AUT). This makes the testing process faster, since each machine can run its own set of tests concurrently.

- **Avoidance of redundancy** — Consider that a web application behaves differently with users with different permissions. It is important to have one configuration for each type of user, to be able to test all the features. However, having more than one configuration for each type of user will lead to redundancy, as these extra tests will not add anything different to the Test Suite. They will not find new errors. By limiting the configurations and applying random testing, the same test results can be obtained with a smaller test suite.
- **Easier debugging** — Since each Test Path is stored into a different XML file, if a critical error occurs that may stop the execution of the testing program altogether, it is easier to see which element or configuration caused that error. Also, as the results are updated after each Test Path, there is no need to run all the selected tests again.

6 Conclusions and Future work

This paper proposed a test generation and filtering technique for mode-based testing of Web applications. The models contain information in the form of UI Test Patterns linked with connectors. Each UI Test Pattern contains specific configurations, obtained during users execution traces.

Several filters were applied to provide flexibility and reduce the number of test

cases generated. Firstly the filters were applied to the Test Paths and then to each Test Path element (an UI Test Pattern) configurations.
This approach proved to be effective, as the filtering can provide the same results with a smaller number of test cases.
The features planned for the near future include a broader set of filtering options and test coverage statistics.

Acknowledgments

This work is financed by the ERDF — European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT — Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-020554.

References

1. Dalal, S. R., Jain, A., Karunanithi, N., Leaton, J. M., Lott, C. M., Patton, G. C., and Horowitz, B. M.. 1999. Model-based testing in practice. In Proceedings of the 21st international conference on Software engineering (ICSE '99). ACM, New York, NY, USA, 285—294.
2. Dick, J., Faivre, A., 1993. Automating the generation and sequencing of test cases from model-based specifications. In: Woodcock, J., Larsen, P.G. (Eds.), FME 93: Industrial-Strength Formal Methods, LNCS 670, pp. 268—284. Springer.
3. Katsiri, E., Mycroft, A. "Model checking for sentient computing: an axiomatic approach." Proceedings of the First International Workshop on Managing Context Information in Mobile and Pervasive Environments (SME 2005), CEUR-WS, Ayia Napa (May 2005). 2005.
4. Claessen, K., Hughes, J., 2000. QuickCheck: a lightweight tool for random testing of haskell programs. In: Odersky, M., Wadler, P. (Eds.), Proc. of the 5th ACM SIGPLAN International Conference on Functional Programming (ICFP 00), ACM. pp. 268—279.
5. Mealy, G. H. (September 1955). "A Method for Synthesizing Sequential Circuits". Bell Systems Technical Journal 34: 1045—1079
6. Kwang-Ting Cheng; Krishnakumar, A. S., "Automatic Functional Test Generation Using The Extended Finite State Machine Model," Design Automation, 1993. 30th Conference on , pp.86-91, 14—18 June 1993
7. Frohlich, P., Link, J. (2000). Automated Test Case Generation from Dynamic Models. In ECOOP 2000, LNCS 1850 (pp. 472—491). Springer.
8. Rayadurgam, S.; Heimdahl, M. P. E. (2001). Coverage based test-case generation using model checkers. Proceedings. Eighth Annual IEEE International Conference and Workshop On the Engineering of Computer Based Systems-ECBS, 83—91.
9. Tretmans, J., 1996. Test Generation with Inputs, Outputs, and Repetitive Quiescence. Software-Concepts and Tools 17, 103—120.
10. Phalippou, M.: Relations d Implantation et Hypotheses de Test sur des Automates a Entrees et Sorties. PhD thesis, L Universite de Bordeaux I, France (1994)

11. Jard, C., Jeron, T.: TGV: Theory, Principles and Algorithms: A Tool for the Automatic Synthesis of Conformance Test Cases for Non-Deterministic Reactive Systems. Software Tools for Technology Transfer 7(4) (2005) 297—315
12. Hartman, A., Nagin, K.: The AGEDIS Tools for Model Based Testing. In: Int. Symposium on Software Testing and Analysis - ISSTA 2004, New York, USA, ACM Press (2004) 129—132
13. He, J., Turner, K.: Protocol-Inspired Hardware Testing. In Csopaki, G., Dibuz, S., Tarnay, K., eds.: Int. Workshop on Testing of Communicating Systems 12, Kluwer Academic Publishers (1999) 131—147
14. de Alfaro, L., Henzinger, T. A., 2001. Interface automata. In: ESEC / SIGSOFT FSE, pp. 109—120.
15. Campbell, C., Grieskamp, W., Nachmanson, L., Schulte, W., Tillmann, N., Veanes, M., 2005. Testing concurrent object-oriented systems with spec explorer. In: Fitzgerald, J., Hayes, I.J., Tarlecki, A. (Eds.), FM, Springer. pp. 542—547.
16. Feijs, L. M. G., Goga, N., Mauw, S., Tretmans, J., 2002. Test selection, trace distance and heuristics. Proceedings of the IFIP 14th International Conference on Testing Communicating Systems (TestCom 2002), pp. 267—282. Kluwer.
17. Nachmanson, L., Veanes, M., Schulte, W., Tillmann, N., Grieskamp, W., 2004. Optimal strategies for testing nondeterministic systems. In: Avrunin, G. S., Rothermel, G. (Eds.), Proc. of the 2004 ACM SIGSOFT International Sym- posium on Software Testing and Analysis (ISSTA 04), pp. 55—64.
18. Moreira, R., Paiva, A., Memon, A., "A Pattern-Based Approach for GUI Modeling and Testing". Proceedings of the 24th annual International Symposium on Software Reliability Engineering (ISSRE 2013), 2013
19. Welie, M., Gerrit, C., Eliens, A., "Patterns as tools for user interface design". Workshop on Tools for Working With Guidelines. Biarritz, France, 2000
20. J. J. Garrett. "Ajax: a new approach to Web applications". http://www.adaptivepath.com/publications/essays/archives/000385.php .2006
21. L. L. Constantine, L.A.D Lockwood. "Usage-centered engineering for Web applications". IEEE Software Journal, vol.19, no.2, pp.42—50, Mar/Apr 2002
22. Monteiro, T., Paiva, A., Pattern Based GUI Testing Modeling Environment in 4th International Workshop on Testing Techniques & Experimentation Benchmarks for Event-Driven Software (TESTBEDS 2013), 2013
23. Nabuco, M., Paiva, A., Camacho, R., Faria, J., Inferring UI Patterns with Inductive Logic Programming in 8th Iberian Conference on Information Systems and Technologies, pp.-, 2013
24. Cunha, M., Paiva, A., Ferreira, H., Abreu, R., PETTool: A Pattern-Based GUI Testing Tool, in 2nd International Conference on Software Technology and Engineering (ICSTE'10), pp.202-206, 2010

Information Systems Security Audit – A real life scenario and prototype solution

Pedro Emanuel Cardoso de Sousa

Department of Informatics Engineering, Faculty of Engineering,
University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto,
Portugal

Abstract. This paper is based on a case study scenario where a major data breach happens, in one institution of public sector, a municipality, from Portugal. The focus of this paper is to show a real life data breach scenario and the steps to solve the problems, not a global solution for all situations, but a solution for this particularly problem. After the real life problem we show a prototype of one solution to increase security in ERP systems, by colleting relevant data to correlate and math against segregation of duties policies.

Keywords: SoD, Security, Information Security, Segregation of Duties

1 Introduction

From the point of view of the organization's management, computer security is seen as an irrelevant factor in the business concerns. This area is usually a concern for the departments of information and communication technologies. This minimizes the factor and the scope of computer security being reviewed at the local level of the department and not as a general concern. This idea is not true, generating a false sense of security for the organization. This feeling lasts while the organization is not compromised. From the start, that business management platform, known as Enterprise Resource Planning (ERP), is the privileged place for the storage of important information. These are careless in terms of computer security, and investments/measures in security are targeted to the equipment and services of perimeter systems, such as firewalls, intrusion detection systems, proxies and security devices. There is a gap between the development and implementation of ERP and computer security efforts made by organizations, the investment area focused in other areas forgetting ERP systems. On the other hand, software development companies have little focus on the issue of computer security and the development of systems is careless in these themes. In a society where all are connected to the network and the internet, information systems security or computer security is not a major concern for organizations in general, in Portugal. This reality has changed in the last few years, but organizations are not fully alerted to this problem and IT (Information and Technology) Departments need a cultural change and the mindset needs to change. Security problems in the public sector need to be a major concern, because public institutions have information from citizens and are financed by taxing money, the money of every contributor. The reality is very different, normally security issues are the last concern for IT Departments in the public sector, and in this paper we have a case study about a real life situation of how security problems can change one organization. This paper intends as follows: Section 2 is the literature review - we defined to take a closer look of various fields in computer security or information security and this look takes a tour into technology and formal methodology; Section 3 describes a real world case study based on real security incidents and we take a look at the improvement that can be done in an organization if we have concerns about security; Section 4 is the future research directions chapter where all key points inside the work is focused and what can be done to improve in the future. The conclusion of this proposal and the next steps to be taken are presented in Section 5.

2 Literature Review

The security problem of ERP starts before implementing in the organizations, ie, starts in the development phase and in software development companies. Traditionally companies that produce software don't have security concerns, in the initial stages of development, and don't have the human resources with expertise in this field. There are some concerns in terms of users and passwords and the need to create a model for privileges inside the software, but the focus of professional in software development are the features and purpose of the operation of the product, according to customer specifications or market where the software is located. There is a gap between software development and the field of computer security and information security. We found highly specialized human resources in software development and highly specialized human resources in computer security. The problem lies in the lack of cooperation and communication between these two fields in the area of information and technology. VAN WYK & MCGRAW (2005). The software development companies wake up to this issue only after the implementation of software to their customers and when security incidents happen. Creating a culture of interdisciplinary work between software development and computer security can lead to the development of a top quality product and avoid many future problems in the implementation of systems. It is not only necessary to implement an interdisciplinary culture between the two fields, it is necessary to create models for the various specialists to communicate and understand the concerns and both parties, because the human resources software development does not dominate the themes of information security and information security specialists are not programmers or developers. You can find security methodologies to software development, the same way we found the UML (Unified Modeling Language) or other languages for communication between the client, analysts and programmers. VAN WYK & MCGRAW (2005). The safety of an ERP system is a continuous process, according to Marnewick & Labuschagne (2006), this process begins with the initial design of ERP implementation and ends with the entry into full operation. However, the safety assessment cannot have this vision, as new technologies emerge every day, as well as new threats and the need to keep the information intact, is a constant and daily goal. Moreover, the ERP system is an integral part of this organization and adapts and changes daily, security must accompany all internal and external changes that occur within the purview of the organization. According to Alexander & Ayers (2011), the continuous monitoring and control is the best weapon to approach zero the possibility of a security incident happens. However there is a need for systematic procedures for monitoring and control, in order to implement a framework that will ensure the safety of the ERP system. Risks exist, but generally are not identified and enumerated in order to understand whether they are controlled. The most common risk is fraud, deliberate or unconsciously, causing data corruption and that has the effect of generating false information. The SIEM (Security Information and Event Management) appear as a mechanism for excellence to support the collection and correlation of security logs, ensuring the traceability of operations, evaluation of the dangers and implementing alerts on activity systems. However, these systems are focused on traditional systems (firewall, servers, authentication, etc.), because the job requires ERPs linked to a technological vision, but also organizational. Lane (2010).

3 Real Life Problem in Municipality Sector

3.1 Real Scenario

The case study scenario is conducted in the municipality of Portuguese territory, a public sector institution and applies to ERP that supports the organization's business.

Information system data for case study:

- Number of staff that works with ERP: > 200 persons;
- Database server operating system: Linux CentOS 5.x;
- Database server system: Oracle Database 10G R2;
- Application server operating system: Windows Server 2003 Standard Edition;
- Application system: SAGA System from Medidata;
- Domain controller: Yes, there are 2 domain controllers supported by Linux;
- Computers are integrated into the domain: All computers are integrated;
- Audit system: The ERP has no audit system;
- Activity logs: The ERP has no activity logs;

How the information system works:

Databases: The system has one instance, called MC, and each application inside the ERP has a specific owner where the tables, views and procedures are stored. For example, the application of construction has an owner called OBP and accounting application has an owner called CTA2011 (or CTA2010, or CTA2012), in this case there is one owner for each year; **Privileges system:** The system security privileges are developed by the supplier of ERP and controlled based on tables within the database; **Database connection:** The connections to the database are made by ODBC. Users and passwords are stored in a local file, along with the executable for each application. Users used for binding, are the owners of the database tables;

The next steps describe the process of connecting to the ERP:

- Step 1: Each user accesses the file server to run the application. This access is guaranteed by a shortcut installed on each computer on the network;
- Step 2: The application is loaded with data from the local file, along with the executable, where users are stored and passwords connected.
- Step 3: On the computer user is a connection over ODBC (Open Database Connectivity), to Oracle database with data from the local file;
- Step 4: The database connection is established with full privileges because the user owns the connection tables;
- Step 5: The application is started and you can work; operations are controlled by the security privileges of the application;

3.2 Security Problems

In the scenario presented, some computer security issues are raised. These issues are placed at various levels, ability to audit, control, access and logging. One major issue is the lack of activity records in the database and one audit system for cases where security incidents happen. ERP systems, especially the system with financial functions, should have mechanisms to register the activity performed to enable auditing in the event of an incident. No system is one hundred percent a proof against security incidents, for this reason software development companies need to implement independent mechanisms for activity records. These records allow access to all activities inside database and leave forensic evidence for security auditors. Another big issue, in this scenario is the access to the database. All users connected to the infrastructure of the municipality have direct access to three key points in the ERP. **Key point 1:** All workstations inside the municipality network have access to the applications executable, because these files are stored in a file server, and the user only needs one domain account to have access; **Key point 2:** Inside the file server, in the same directory as the application executable, we have the local file with the data for connecting the database. This file has the username and password to connect to ERP database; **Key point 3:** All devices inside the infrastructure have one ODBC connection to the database. To have one SQL (Structured Query Language) console, the user only needs the username and password, and this information is inside the local file, stored together with the application executable on the file server.

Another issue in this institution is the lack of records for remote connections to the infrastructure. In the network of the municipality we have remote buildings with fibre optics connections, remote buildings with Microsoft Remote Desktop Connections (MS-RDP) and users with Virtual Private Network (VPN) connections. In the IT (Information and Technology) Department there isn't any records for these remote connections and the global IT infrastructure don't have one centralized logging system to control and monitor the activity from remote connections. The final issue in this scenario is the root and administrator accounts. These accounts are shared by users of the

IT Department, outsourcing human resources and suppliers. In case of one security incident it's impossible to isolate the access to core systems and functions, because many people know the root and administrator access. These accounts are responsible for managing the infrastructure.

3.3 Problems in the ERP System

In this municipality two major security incidents happened in the past related to the ERP and core functions for the organization.

First problem: The security incident was located in ERP and in their servers without affecting other software, not the normal operation of other services within the network of the municipality. Even other computer systems that were within the same database server, but in different instances, were not affected by this security incident. With the scenario presented in this municipality, it was impossible to identify the origins of the incident due to lack of records and liberalized access to the computer system without any control. The only variable was the common focus on ERP supplier Medidata which was used by more than 80% of human resources of the municipality. Another factor was the randomness of the occurrence of the abnormality of the system, i.e. the system was retrieved based on a policy backup of 30 to 30 minutes but the abnormalities computer hour took place in completely separate within working hours. This problem was an additional factor, which increased the complexity of forensic analysis, remote access to the information system. The origin of security incidents could be inside the network of the city, one of the remote buildings or any of the clients who had access VPN. These remotes were not recorded or monitored.

A step to solving: After the analysis of the technological infrastructure of the municipality and the lack of control of key points in the information system and computer network, we began a process of reconfiguration of the infrastructure to isolate the source of the problem and ensure the existence of records and evidence in case security incident. We have performed several operations: Revoke all remote access to the network of the municipality; Encrypt files from local users and passwords; Increase the frequency of backups to minimize the impact of data destruction; Change the passwords of users (root and administrator) of systems administration from infrastructure, ensuring that only administrators had knowledge; Remove all ERP applications for administration, who were on the network shares that users had access; Ensure that no system administration task would cause a security incident;

The problem of software development without security concerns.

In the scenario presented is an instance of the database within an Oracle database, where all ERP has its applications. This ERP comprises several applications and application modules and each has a table structure itself. As previously stated, the division of these tables is done with the structure that Oracle provides owners, or each application has an owner in the database and those are within this owner tables and views the application operation. The problem with this approach is the connection method, ie, is the owner of the tables used to connect to the database; it is not possible to restrict the privileges of the user before the database objects, because he owns the objects. The operations that were to be performed on the database were DROP TABLE and DROP VIEW, destroying the structure of the database without leaving operation records. By destroying the structure, data were also deleted, causing the collapse of the information system. Analyzing in detail, the user only requires access to one of the network devices in the municipality to obtain a SQL console to the database and cause the shutdown of the system.

Change system security

After analyzing the operation method of the ERP software that supports the municipality, it was possible to increase the security level of the data layer with changes in the structure of the database. Example of behavior of an ERP application:

- Application: OBP
- Owner: OBP
- Connection user: OBP
- Table: OBP.TABLE_A
- View: OBP.VIEW_A
- Call method (path): SELECT * FROM TABLE_A

After analyzing application behavior, we learn that all objects (tables and views) are called by the local name, i.e., the user connection is always the owner of the objects and these are never called by full path, only for short name or short path. With this analysis, we create the structure of the database, ie by each owner (owner A) (user) from database, we will create a new user (owner B) with the suffix _USER. Then we will ensure that the new user can only perform the four operations to the old SELECT, UPDATE, INSERT and DELETE. After removing the ability to connect with the owner (owner A) of the tables and added a synonym for each object from owner A to the owner B, simulating the local existence of tables and views.

Example of behaviour of an ERP application:

- Application: OBP
- Owner: OBP
- Connection user. OBP_USER
- Table: OBP.TABLE_A
- View: OBP.VIEW_A
- Synonymous: OBP.TABLE_A -> OBP_USER.TABLE_A
- Synonymous: OBP.VIEW_A -> OBP_USER.VIEW_A
- Call method (path): SELECT * FROM TABLE_A
- Operations between both: SELECT, UPDATE, DELETE and INSERT

Following this transaction in the database, we changed the local file where the data connection to the application are stored, and put the information for the new database users. With this change, we guarantee that the applications can only perform basic operations over the data in the database (SELECT, UPDATE, INSERT and DELETE).

Second problem: After the improvements made in the previous paragraph the information system found a point of stability and finished security incidents that caused the collapse of the system and stopped the services. The improvements applied to the information system were designed to stabilize the system information and then identify the source of the problem. After stabilizing the system, there were no more anomalies or security incidents, therefore it was not possible to identify the source of the problem. After some time of normal system operation, reoccurred anomalies and security incidents, but now with a different pattern. With the restrictions described in the preceding paragraph, it was not possible to destroy the information system, however this has entered a new cycle of problems but located in two ERP applications, application OBP (processes of construction works) and CTA (accounting processes and financial movements). The fact that the incidents are located allowed to isolate the problem and conducted to a more detailed audit.

The problem in the system: Some system tables were subject to DELETE operations, eliminating all the records inserted. These tables were system tables and do not exist in the application functions or options that allowed a user to manipulate existing data. The elimination of these data caused the shutdown of the system because they were essential to the creation and manipulation of active processes in ERP. Despite the anomaly is found, the system does not hold any registration mechanism that allows analyzing and auditing functions or users who were manipulating the data that were deleted. Nor was it possible to associate with the users of ERP because they were based on the application structure itself without any connection to objects in the database. Data transactions are recorded by the system database and it was necessary to create a mechanism to register for these operations their audit logs, it was possible to analyses operations and identifies the cause of the problem. From the analysis of the entire information system, described over the various points of this document, we know that we cannot rely on the analysis of ERP users; we also know that all equipment is in the area of the county and that each person has a domain user only. Within the database, we could know the user connection, the table involved, the operations performed and the data manipulated. With all this information, a system was designed to record necessary data, to correlate the operations performed within the database, the domain user and the device.

We created a database of independent ERP where all records of data manipulation in the system were sent. The structure of the log table aims to cross the operations performed on the database with the source outside the network, the execution of operations. After the creation of this system, it was added a Trigger on each table allowing the ERP to record this data for any operation performed on the tables. Not to cause delay or influence the normal operation of ERP and its applications, it was chosen non-intrusive method for the execution of operations, so the system works silently without jeopardizing the work of users and making it impossible to be detected by the attacker, if it is a person. After several days of monitoring and various anomalies caused by security incidents, we audited records of monitoring and auditing system, where it was possible to perform a forensic analysis and identify how the security incident was caused by a computer user. After crossing with the attendance system and collecting testimonials from other human resources of the municipality confirmed the presence in the place and time of the attacker thus triggering the legal process for such situations.

4 Prototype Solution

4.1 Test Scenario

For this project we will use a very popular ERP system in the Portuguese market, Primavera Software, using the free version of the solution. With this solution we simulate processes in an organization and validate the collection of information and the application of models of segregation of duties over the activity performed in the system. The whole scenario will be validated in a simulated environment, because due to the time and scope of the project is addressed only a small part of the theme, in order to prove the concept and make a bridge to future works.

4.2 SoD Policies to test in the scenario

Boundary method for applications: This policy aims to validate that applications can consume the data that are in ERP , limiting the use of third party applications, with the goal of always validate that a strange system is the application to manipulate data from our ERP . These rules will be extended to the pair application / computer. To validate the rules are described below:

- The base implementation of ERP has no limit boundary, can consume data from any location.
- The management application database is limited to the administration computer on the network defined.
- All other applications cannot consume the data.

Boundary method for users: This policy aims to limit the border of manipulating tables through the functions of the department in which it operates. The clear example is the department of human and financial resources, ie , most of the time the tables are handled the same and the human resources department should not manipulate data from others. When this happens we are dealing with a breach of boundary method of user functions. To validate the rules are described below: The user A can only handle a given set of tables in the case of articles table. User B can only handle a certain set of tables in case the customer table.

Method validation process and separation of powers: This is one of the most complex methods, ie , we create stages for a given process and separate functions limiting the actions of the user A and B. User A can only execute the launch of the invoice and user B is that it has a responsibility to send payment .

4.3 Technological scenario

The technological landscape is composed of the following elements:

- Database Server with Microsoft SQL Server and is where the information is stored.
- Operator Station will be installed where the management application, if the ERP;
- Tour management that simulates the functions of the system administrator and a second operator.
- Server database auditing platform.
- Web server interface for the management of the audit platform.
- Collection of information agents, installed on each data source.

4.4 Operation

Phase 1: Configuring the database - This step is an analysis of the database to monitor and configure the respective accessions;
Phase 2: Creation of dictionaries - This step consists of research and collection of data sources within the database to monitor and creating a dictionary that represents the knowledge of the database for our auditor system;
Phase 3: Installing monitoring agents - Of the dictionaries created us define which data sources that we monitor and therefore we will install an agent on each font set;
Phase 4: System Settings - In this phase the policies are configured, users, rules, etc. for our auditor system validate the data collected and act accordingly;
Phase 5: Data Collection - From this weighting can start the collection of data and monitoring of our system.

4.5 Collected data - Examples

The system consists of 2 large data repositories that relate to transactions and data collected. Repository for transactions:

- [id_transaction] [uniqueidentifier] NOT NULL ,
- [session_id] [bigint] NOT NULL ,
- [net_transport] [nvarchar] (255) NOT NULL ,
- [encrypt_option] [nvarchar] (255) NOT NULL ,
- [auth_scheme] [nvarchar] (255) NOT NULL ,
- [host_name] [nvarchar] (255) NOT NULL ,
- [program_name] [nvarchar] (255) NOT NULL ,
- [client_interface_name] [nvarchar] (255) NOT NULL ,
- [login_name] [nvarchar] (255) NOT NULL ,
- [nt_domain] [nvarchar] (255) NOT NULL ,
- [nt_user_name] [nvarchar] (255) NOT NULL ,

- [original_login_name] [nvarchar] (255) NOT NULL ,
- [connect_time] [datetime] NOT NULL ,
- [login_time] [datetime] NOT NULL ,
- [audit_table] [nvarchar] (255) NOT NULL ,
- [audit_command] [nvarchar] (255) NOT NULL ,
- [audit_trigger] [nvarchar] (255) NOT NULL ,
- [audit_process] [nvarchar] (255) NOT NULL ,
- [op_time] [datetime] NOT NULL ,
- [db_name] [nvarchar] (50) NOT NULL ,

The fields above presented aims to collect data on the source transaction and the system you are using, to ensure trace back to the origin of the transaction.
Data repository:

- [op_id] [bigint] IDENTITY (1,1) NOT NULL ,
- [table_name] [nvarchar] (255) NOT NULL ,
- [field_name] [nvarchar] (255) NOT NULL ,
- [value_new] [nvarchar] (4000) NULL ,
- [value_old] [nvarchar] (4000) NULL ,
- [operation] [nvarchar] (255) NOT NULL ,
- [op_time] [datetime] NOT NULL ,
- [transaction_id] [uniqueidentifier] NOT NULL ,

5 Future Research Directions

The software security is a field of information and technology that connects software development and computer security, this is a growing field and hard work to develop. The study of the problem of security information at the source, i.e. producing software in order to increase the final quality of the product development is a focal point for the new generation of software, especially with the generation of cloud computing closer, the home user. The search methodologies and standards that define best practices for companies to follow software development will be a differentiating factor for professionals. The education sector will also have to evolve in this direction, so that professionals are aware of the concerns of the market and what paths they can choose to find solutions to this problem. Typically the management persons are not alerted towards the security problems and their knowledge about computer security is based on common sense. These concerns arise when the first security incidents in the organizations they manage. Even if the management is alerted to the problems of computer security is a big gap between technical knowledge and concerns. The implementation of methodologies such as segregation of duties enables security specialist communicate with the design and management of organizations audit and control processes that have a focus on the real concerns of the organizations. This is a field of study course, since the establishment of formal methodologies that allow communicating with decision elements that have no technical knowledge is essential. The next step in the creation of formal models, for enabling communication with people management organizations, is creating methodologies for implementation and execution. You cannot create a model for control and audit information security, such as segregation of duties, if it has not at the level of implementation. The ability to execute the model in real and practical cases is critical to success and growth. The study of mechanisms for implementing these models is a growing field of computer security. Monitoring systems, control and audit information systems are an essential tool, even if there is already on the market something that allows multiple systems to collect and correlate logs, these are not oriented information systems, particularly for ERP. The study of the adaptation of existing methodologies and mechanisms for the collection and correlation of data from network equipment and its adaptation to the information systems (ERP) is still an unexplored field. Another factor that is important is the ability to treat large volumes of data collection systems have the ability to timely analyses and detect anomalies within defined policies and anomalies that escape the normal functioning of the systems. Fields of study such as artificial intelligence are the future (and present) of data correlation and in particular in supporting the identification of problems or anomalies. These are a clear growth in the market and are many perspectives of development, however there are several prospects for developing direct and concrete relationship to this work. Firstly the closure of features in a real scenario, and the test work a general organization. Secondly validate the performance issues in large where the collection of these data can be superior to own ERP systems. Thirdly generalize functions and procedures to try to create a system capable of covering several solutions and correlate data from multiple platforms. Fourth adding intelligence capability not only based on a static model rules, in particular detection method uses standard off. Finally a more detailed analysis of some Open Source and Freeware solutions that are on the market, with the implementation of own developments for smaller architectures and lowest volume of information.

6 Conclusion

This work is based on a practical approach to security issues in information systems, in particular ERP systems in the public sector in Portugal. The scenario presented, the problems identified the solutions adopted and the paths taken are based on a real case study. The problems presented are common to many organizations that have the feeling that the information system is safe, even if they have a lot of money invested in technology infrastructure and information system. The problems of computer security should be constant concern, which must be analyzed in terms of management of the organization and operated by IT departments and suppliers. The management level is important because it is in this dimension that we can understand what areas and processes are important for the organization, so that investments are invested in the areas of information systems that are important to business. The case study presented also indicates that huge investments are not generally required to increase the security level of an organization. To understand this dimension, it is necessary to analyses the information system as a whole and with concerns in the field of computer security. Bearing in mind this, we will see where and how we should improve the information system and the organization's policies. The involvement of technology providers is a key to success, but must be available to effect changes in the systems they sell. On the other hand, we realized that it is necessary to perform a job structuring at the level of culture of software houses and the internal staff, normally the developers. This work should be initiated by the education system, because the new generation of developers should be awoken to these concerns and the needs of organizations. Education institutions should look at these concerns as a challenge and not as a problem, because there is a window of opportunity to increase the quality of professionals, if synergies of collaborative work between programmers and computer security experts are converged. In the prototype it was possible to prove the generalization of concepts and non-intrusively monitor a very popular ERP in the Portuguese market. Although the solution found to be quite limited by the size of the prototype and the simulation environment, it is possible to realize the growth potential. Of course this requires prototype testing on real and large volumes of information scenarios.

7 References

1. Alexander, C., & Ayers, A.,2011. ERP security & continuous controls monitoring (CCM). RSM Tenon.
2. Goel, S., Kiran, R., & Garg, D., 2012. Vulnerability management for an enterprise resource planning system. International Journal of Computer Applications, 19-22.
3. Hertenberger, M., 2005. Issues in information systems issues in information systems. Johannesburg: university of Johannesburg.
5. Lane, A., 2012, April 30. Understanding and selecting SIEM/LM: Use cases. retrieved 11 01, 2012, from Securosis Blog: http://securosis.com/blog/understanding-and-selecting-siem-lm-use-cases-part-1
6. Marnewick, C., & Labuschagne, L., 2006. A security framework for an ERP system. Auckland Park, South Africa: University of Johannesburg.
7. Saharia, A., Koch, B., & Tucker, R., 2008. ERP systems and internal audit. In Issues in Information Systems, 578-586.
8. Swart, R., Marshall, B., Olsen, D., & Erbacher, R., 2007. ERP II System Vulnerabilities and Threats: An exploratory Study. IRMA International Conference, Idea Group Inc, 925-929.
9. Tie, W., & Cheng, W., 2011. Study on Enterprise Information Security in the ERP Conditions. International Conference on Computer Science and Network Technology, IEEE, 332-335.
10. VAN WYK, K., & MCGRAW, G. (2005). Bridging the Gap between Software Development and Information Security. IEEE SECURITY & PRIVACY, 75-79.
11. Wolf, P., & Gehrke, N., 2009. Continuous compliance monitoring in ERP systems – A method for identifying segregation of duties conflicts. Hamburg: Wirtschaftinformatik.
12. Yu, X., & Wen, Q. N.d.. A View about cloud data security from data lifecycle. Beijing, China: IEEE.

Support for automated teamwork submission leading to automated plagiarism detection

Rodolfo Matos

Shared Services – SPUP, University of Porto – U.Porto, Porto, Portugal
rodolfo@sp.up.pt

Abstract. The submission of work by groups of students tends to entail a lengthy process of administration and control of the submissions made. With the massification of education and the increasing focus on peer education and teamwork, the administration tasks of the submission process should be reduced to a minimum. Otherwise, the time spent in management tasks, together with the time required to evaluate the submissions, can make the whole process impractical. This article proposes an automated submission web site in order to automate access to submissions in an organized fashion that in turn make it possible to use automated plagiarism detection tools. Implementation details are shown of a system that was developed precisely to address this problem.

Keywords: Deliverable submissions, management, peer education, teamwork, Course Portfolio

1 Introduction

Managing deliverables submission can take an unsuspecting large amount of time. However, due to the massification of education programs, team works are becoming increasingly frequent. The combination of these two factors, may incur in some very demanding managing tasks, making evaluators spend much more time managing the system than they initially intended.

In a previous work [1], we used the course "Projeto FEUP" [2] as a case study for a semi-automated way of producing an organized set of all deliverables from a given course in a given year. Most managing tasks on that work came from fixing wrong deliverables filenames, fixing conflicting files sent by different persons from the same group, the submission of different versions of the same document and several other similar situations.

Since the course had about 1056 students, splitted over 180 groups, each group having to submit four different deliverables, notwithstanding the seemingly tightest controlled environment, there was space for errors, although most objectives were achieved.

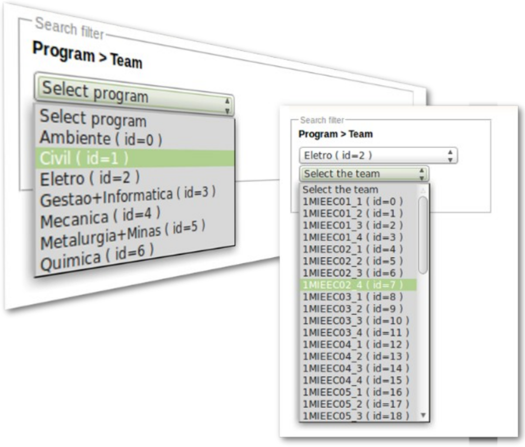


Fig. 1. "Projeto FEUP" program teams

This work presents another approach, which proved to be much more flexible and efficient. We opted to use a loose control approach, on which the students could themselves control each other's submissions. Several ways of checking automatically the similarities between documents were tried. The final prototype used Sherlock [3] to generate a report by evaluators' request. This approach released the evaluators from most of the management tasks, exactly as we were expecting.

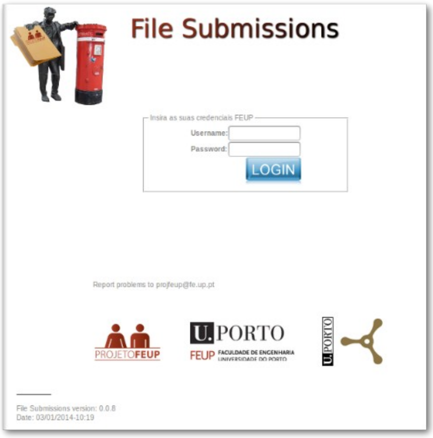
2 State of the art

Of the various methods of submitting deliverables from groups of students, seldom if ever are the ones that use a loose control approach towards the students' submissions. Although there are a lot of interest on a portfolio usage, the actual process of submitting deliverables is often bound to have a controlled environment. This normally means that either the name of the files and/or location are fixed by evaluators. As an immediate consequence of this follow-through procedure, checking if a file submission is in proper conditions is not done automatically, there is a good probability that some of the students have problems.

3 System Overview

The process that we used in order to reduce the management needs to a bare minimum, was "social pressure by peers to get things right". This meant, at its core, that every authenticated student could upload whatever files he wanted. This apparently insecure approach, proved to be extremely efficient and secure, essentially due to three factors:

1) A Shibboleth/LDAP authentication was implemented, allowing only properly authenticated users to use the system



File Submissions

Insira as suas credenciais FEUP

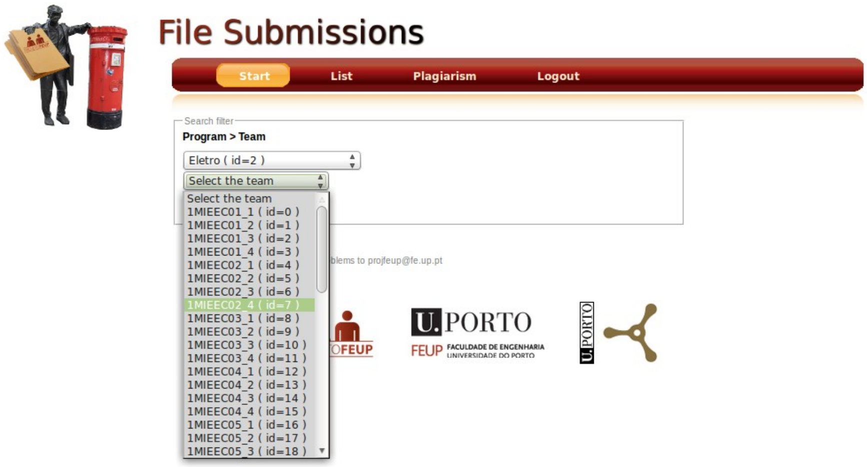
Username:

Password:

Report problems to projfeup@fe.up.pt

File Submissions version: 0.0.8
Date: 03/01/2014-10:19

Fig. 1. Authentication page



File Submissions

Search filter

Program > Team

Eletro (id=2)

Select the team

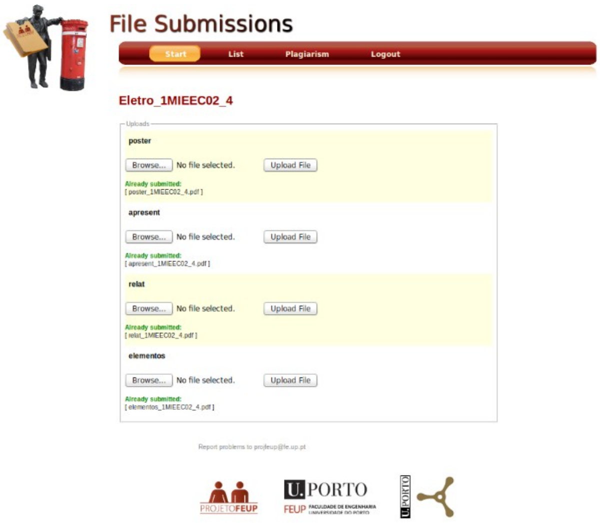
Select the team
1MIEEC01_1 (id=0)
1MIEEC01_2 (id=1)
1MIEEC01_3 (id=2)
1MIEEC01_4 (id=3)
1MIEEC02_1 (id=4)
1MIEEC02_2 (id=5)
1MIEEC02_3 (id=6)
1MIEEC02_4 (id=7)
1MIEEC03_1 (id=8)
1MIEEC03_2 (id=9)
1MIEEC03_3 (id=10)
1MIEEC03_4 (id=11)
1MIEEC04_1 (id=12)
1MIEEC04_2 (id=13)
1MIEEC04_3 (id=14)
1MIEEC04_4 (id=15)
1MIEEC05_1 (id=16)
1MIEEC05_2 (id=17)
1MIEEC05_3 (id=18)

blems to projfeup@fe.up.pt

Fig. 1. Program team selection

2) A logging system was deployed, making all actions carried out by students readily available to system managers, thus making it easy to identify and correct anomalous situations

3) The deliverables list was visible to entire community, thus making the process transparent and allowing others to check “who sent what where” naturally built a peer pressure that helped to regulate and prevent abnormal situations with minimum effort.



Eletro_1MIEEC02_4

Uploads

poster

Browse... No file selected. Upload File

Already submitted:
[poster_1MIEEC02_4.pdf]

apresent

Browse... No file selected. Upload File

Already submitted:
[apresent_1MIEEC02_4.pdf]

relat

Browse... No file selected. Upload File

Already submitted:
[relat_1MIEEC02_4.pdf]

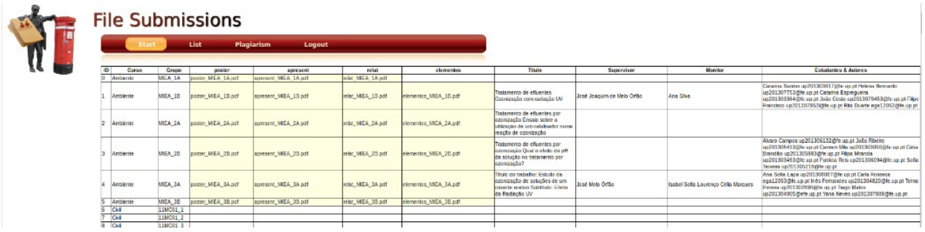
elementos

Browse... No file selected. Upload File

Already submitted:
[elementos_1MIEEC02_4.pdf]

Report problems to projfeup@fe.up.pt

Fig. 1. Team submissions page



ID	Curso	Grupo	poster	apresent	relat	elementos	Título	Supervisor	Monitor	Estudantes & Autores
0	Ambiente	MIEA_1A	poster_MIEA_1A.pdf	apresent_MIEA_1A.pdf	relat_MIEA_1A.pdf					
1	Ambiente	MIEA_1B	poster_MIEA_1B.pdf	apresent_MIEA_1B.pdf	relat_MIEA_1B.pdf	elementos_MIEA_1B.pdf	Tratamento de efluentes Ozonização com radiação UV	José Joaquim de Melo Órfão	Ana Silva	Catarina Santos up201303617@fe.up.pt Helena Bernardo up201307753@fe.up.pt Catarina Espregueira up201303364@fe.up.pt João Costa up201307945@fe.up.pt Filipe Francisco up201307953@fe.up.pt Rita Duarte ega12062@fe.up.pt
2	Ambiente	MIEA_2A	poster_MIEA_2A.pdf	apresent_MIEA_2A.pdf	relat_MIEA_2A.pdf	elementos_MIEA_2A.pdf	Tratamento de efluentes por ozonização Ensaio sobre a utilização de um catalisador numa reação de ozonização			
3	Ambiente	MIEA_2B	poster_MIEA_2B.pdf	apresent_MIEA_2B.pdf	relat_MIEA_2B.pdf	elementos_MIEA_2B.pdf	Tratamento de efluentes por ozonização Qual o efeito do pH da solução no tratamento por ozonização?			Alvaro Campos up201306132@fe.up.pt João Ribeiro up201305413@fe.up.pt Carmen Mila up201303050@fe.up.pt Cátia Brandão up201305993@fe.up.pt Filipa Miranda up201303453@fe.up.pt Patrícia Reis up201306094@fe.up.pt Sofia Teixeira up201305218@fe.up.pt
4	Ambiente	MIEA_3A	poster_MIEA_3A.pdf	apresent_MIEA_3A.pdf	relat_MIEA_3A.pdf	elementos_MIEA_3A.pdf	Título do trabalho: Estudo da ozonização de soluções de um corante reativo Subtítulo: Efeito da Radiação UV	José Melo Órfão	Isabel Sofia Lourenço Célia Marques	Ana Sofia Lapa up201306087@fe.up.pt Carla Fonseca ega12053@fe.up.pt Inês Fernandes up201304820@fe.up.pt Telmo Pereira up201302890@fe.up.pt Tiago Matos up201304005@efe.up.pt Yana Neves up201307809@fe.up.pt
5	Ambiente	MIEA_3B	poster_MIEA_3B.pdf	apresent_MIEA_3B.pdf	relat_MIEA_3B.pdf	elementos_MIEA_3B.pdf				
6	Civil	11MC01_1								
7	Civil	11MC01_2								
8	Civil	11MC01_3								

Fig. 1. File submission list

However, we did implement a control system, which has proved particularly effective, based initially in an open-source plagiarism detection tool called “Sherlock” [3], and a custom-made web interface to act as a dashboard for the tool. The comparison is done between all files submitted “pairwise to find those which are most similar. It uses a digital signature scheme to hashes up N adjacent ‘words’ of input, and semi-randomly throws away many of the hashed values so that it become hard to hide the plagiarized text”. Upon the extraction of the text in the files submitted – mostly PDF's – we were able to check amongst all deliverables, those situations where a duplication of content occurred, be it intentional or by mistake, by finding different versions of the same file in different groups. It was also possible to monitor and assess many other situations we had seen as recurrent even in a tight controlled system [1].

4 Conclusions

By keeping all maintenance and management tasks to a minimum, as was intended, the presented prototype successfully generated the list of submitted files in the course but by reducing enormously the management tasks needed to accomplish it, compared to a more strict approach as was implemented in a previous work.

There was also the deliberate choice of not to use any kind of external database needs on this prototype. The behavior of the system can be tweaked simply by replacing some values in a configuration file and in a CSV (Comma Separated Values) file containing the definition of the organization of the student groups. Thus the deployment tasks – and time to perform those – also reduced as intended.

5 Future Work

This initial prototype was meant to be used on natural language written documents, in which the outcome of the plagiarism tools report would serve only as a thumbs rule to identify problems in the deliverables submissions. So there was no need of increasing the accuracy of the plagiarism detection algorithms used.

Other fingerprinting methods, like winnowing [5] were also considered for future versions of the prototype

The next obvious step is to extend the framework to use local and external repositories of older submissions.

References

1 Sousa A.; Matos R. (2013): Managing Team Work Submissions in Large Moodle Courses in Order to Generate Course Portfolios, CISTI'2013

2 Institutional page of “Projeto FEUP” course at university’s information, https://sigarra.up.pt/feup/pt/UCURR_GERAL.FICHA_UC_VIEW?pv_ocorrencia_id=335490 (accessed Dec 15, 2013)

3 The Sherlock Plagiarism Detector, http://sydney.edu.au/engineering/it/~scilect/sherlock/ (accessed Dec 15, 2013)

4 Adeva J., Carroll N., Calvo R. (2006), Applying Plagiarism Detection to Engineering Education, 7th International Conference on Information Technology Based Higher Education and Training

5 Schleimer S., Wilkerson D., Aiken A., (June 9-12, 2003) Winnowing: Local Algorithms for Document Fingerprinting, SIGMOD 2003

SESSION 4

SIMULATION SYSTEMS & HIGH-PERFORMANCE COMPUTING

CWeaver: An Aspect Oriented C-to-C Compiler

Pedro Pinto

A 3D cross-platform for web-based visualization of spatiotemporal information in real-time

Leonel Dias

Dungeon multi-level generator for a serious game

Ricardo Gonçalves and António Coelho

CWeaver[⋆]

An Aspect Oriented C-to-C Compiler

Pedro Pinto

Faculty of Engineering, University of Porto
pmsp@fe.up.pt

Abstract. When dealing with large programs, manual code instrumentation can be a time-consuming and error-prone task. Additionally, if one intends to apply different instrumentation strategies, it is possible that several versions of the source code are created, reducing the maintainability of our code base. Our hypothesis is that using a source-to-source compiler controlled through an Aspect Oriented approach, can mitigate both problems. First, because it relies on a semi-automated process, we eliminate most of the error-inducing tasks. Second, an Aspect Oriented approach allows us to maintain a single source file separated from the files that describe the instrumentation strategies, merging them as needed. Our solution uses LARA, an Aspect Oriented language, and Cetus, an existing compiler infrastructure, to provide a simple compilation flow that can extract information and create analysis reports or manipulate the original program in order to instrument it. This work presents two instrumentation examples that illustrate the features and advantages of our approach when performing code instrumentation tasks.

1 Introduction

Code instrumentation can be a rather time-consuming and tedious task, depending on the dimension of the code base to instrument. Moreover, if this task is performed manually, it can also be error-prone. Hence, it might be desirable to automate this task as much as possible, while still providing the user with control over its application. The user needs to hold enough information in order to make correct decisions, as certain automatic code instrumentation approaches might change the behavior of the original program.

This work was initially proposed in the context of the AutoSeer [1] project, with the intent of applying sophisticated code instrumentation strategies that allow us to understand how specific code invariants relate to fault location diagnostic precision. For this purpose, we needed a tool that could apply several different instrumentation strategies in a semi-automatic fashion. This activity raises the problem of maintaining different versions of the same source code, one for each strategy to be tested. Aspect Oriented Programming(AOP) [8] can be

[⋆] This work was developed in the context of AutoSeer, a project partially funded by the Portuguese Science Foundation, under grant PTDC/EIA-CCO/116796/2010.

used as an effective approach when dealing with this problem. This paradigm relies on the separation of the primary concerns, the intended behavior of the program, from the secondary concerns, the instrumentation strategies, and on having an *aspect* that manages to successfully describe their relations and how they should be combined.

We propose to develop a tool that is capable of performing source code manipulation, focusing on instrumentation, and that can leverage the advantages offered by an AOP approach. This tool should able to compile C source into C source with the modifications defined in the aspect. We will develop a source-to-source compiler that relies on two proved technologies: LARA [3], a domain specific AOP language that will be used to guide the compilation flow, and Cetus [6], a source-to-source compiler infrastructure for ANSI C. We believe this is a feasible task, capable of producing good results, as indicated by other recent source-to-source approaches guided by LARA (please see Section 6).

The rest of this paper is organized as follows. Section 2 explains the concept of Aspect Oriented Programming and details LARA. Section 3 presents the proposed solution while Section 4 contains a detailed overview of its architecture and main components. In Section 5, we can see two examples that illustrate the main features of CWeaver. Finally, Section 6 discusses related work and Section 7 concludes the paper, while also presenting directions for future work.

2 Aspect Oriented Approach

Although all programming paradigms have their decomposing criteria, there are concerns that do not align with the primary decomposition. These secondary or cross-cutting concerns, are spread along the program decomposition unit, usually repeated and scattered along the source code, polluting the primary objective of the system [7].

Aspect Oriented Programming (AOP) [8] is a paradigm which intends to increase the modularity of programs, by separating cross-cutting concerns from those that describe the intended behavior of the application [7]. This paradigm is based on the idea that a system works better if its properties, non-functional requirements and areas of interest are specified separately and there is an *aspect* describing the relationships between them. The *weaving* process will then bring these concerns together into the intended program [7]. Using aspects to separate secondary concerns from the core objective of the program results in cleaner code, easier concern analysis, monitoring, tracing and debugging.

Currently, there is a diverse set of aspect programming languages, such as AspectJ [9], an aspect-oriented extension for Java, and AspectC++ [13], to apply aspects in C++ programs.

2.1 LARA

The previously mentioned AOP languages define specific aspects, which cannot be reused across different aspect and target languages. For instance, an extensive aspect defined in AspectJ, possibly useful for other languages, cannot be immediately used with other languages without manual translation. Other disadvantages of the mentioned tools include their simple model for points of interest (that, e.g., does not consider loops and local variables) and the reduced number of available actions, usually restricted to code insertion.

LARA [3] is a novel AOP approach that is partially agnostic to the target programming language. It allows the definition of aspects for different target languages, i.e., LARA aspects are generic enough so they can be applied to several problems, over a number of (specified) languages. LARA enables developers to capture non-functional requirements and concerns in the form of strategies, which are completely decoupled from the functional description of the application [3, 4].

LARA differs from other AOP approaches as its syntax is agnostic to the contents of the target language, which are the possible information one can retrieve and the actions one can apply over a program. LARA divides these language specification contents into three components. The *join points*, are points of interest within the code that we intend to observe or influence, e.g., function calls or array accesses. *Attributes*, represent join point information that we can retrieve, for instance, the name of the called function. Activities one can carry out over the selected join point, such as monitoring code insertion, are called *actions*.

With LARA, the language specification is defined externally, as opposed to other common approaches. This allows flexibility for new updates and to easily target different programming languages. As a consequence, LARA does not have a built-in weaving process. Instead, the LARA Compiler parses the aspects and combines them with the language specification to generate the Aspect-IR, an intermediate representation of the aspects for a specific target language. The Aspect-IR should then be interpreted by an external interpreter or weaving tool, such as LARAI [4], the LARA Interpreter. LARA remains partially language-agnostic, as it is the weaving tool that is bound to a specific target language.

3 Proposed Solution

Our solution is CWeaver, a source-to-source compiler for C, that is controlled through an AOP approach, using LARA aspects. This compiler manages to leverage the expressiveness and modularity of LARA aspects, to control the query and manipulation of an Abstract Syntax Tree (AST) of an existing compiler infrastructure, Cetus. This creates an easy compilation process of C source files with the main goal of code instrumentation. The usage of aspects enables an easy selection of points of interest in the code, represented by LARA join points, which can then be analyzed for information retrieval or transformed, through actions. Thus, CWeaver can be used to create information reports based on compiler analyses or to implement complex and sophisticated code instrumentation strategies. CWeaver makes use of two already existing and tested platforms, LARA, the aspect-oriented DSL used to control the compilation flow, and Cetus used as both intermediate representation and back-end component. We believe that using two already proved and tested platforms brings value and robustness to the solution.

CWeaver needs two inputs, the original source code and a LARA aspect. As an example, consider the simple aspect depicted in Fig. 1, that prints to the console the number of iterations each loop is going to execute. The loops are identified by their *rank* attribute, which uniquely identifies a loop inside a source file. This aspect shows the most basic LARA construct, a *select* followed by an *apply*. The former is used to capture points of interest in the source code, loops in this example, and the later is used to act over these points.

```
aspectdef PrintLoopIterations
  select loop end
  apply
    println($loop.rank + ": " + $loop.num_iterations);
  end
end
```

Fig. 1: LARA aspect that prints the number of iterations of each loop.

As a source-to-source compiler, CWeaver is intended to be used as an intermediate tool, part of a more complex toolchain, rather than as a standalone compiler. This usage is perfectly illustrated in [11], where CWeaver is used to transform source code as part of a Design Space Exploration mechanism that aims to increase performance on specific kernels using different multicore models. Despite being in its early stages of development, we believe CWeaver is a flexible source-to-source compiler suitable for a variety of tasks, such as code instrumentation (its main purpose), code transformations and optimizations and even as an alternative to multi-architecture, directive-based compilation.

4 Architecture

CWeaver is composed of three main components, which are presented in Fig. 2. The first component is the LARA Engine, that provides the interface to the user and translates the aspects into commands for the next component. The Weaving Engine is responsible for establishing communication between the other two components. The commands from the LARA Engine are converted into specific actions that are applied over Cetus, which builds and maintains the AST after parsing the input C source files. Each of these components is detailed in the following subsections.

4.1 LARA Engine

The LARA Engine is the component where the execution starts and is itself composed of two subcomponents. The first, is the LARA Compiler, which is

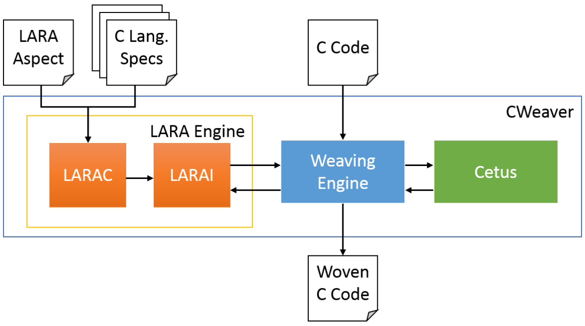


Fig. 2: The architecture of CWeaver.

responsible for compiling the aspect into an intermediate representation, known as Aspect-IR. Because LARA is partially language-agnostic, it also needs as input a description of the target language. This is the *C Lang. Specs* block in Fig. 2, a set of three XML files (provided with CWeaver) that contain the specification of the language: join points, attributes and actions. The generated Aspect-IR is saved to an XML file and passed to the LARA Interpreter, the second subcomponent.

The interpreter takes this new representation and communicates with the Weaving Engine the instructions that it needs to carry, i.e, join point selections, attribute queries or program-altering actions. There is an iterative information flow that goes from the LARA Interpreter, to the Weaving Engine, to Cetus and then back again to the interpreter through the reverse path. The information flows forward in the form of commands to the Weaving Engine and flows backwards in the form of join points or attributes.

The LARA Engine provides the interface with the user, as it is the component responsible for parsing the LARA aspects and initiate the execution, as well as printing the information to the console as in the example of Fig. 1.

4.2 Weaving Engine

The Weaving Engine is the central component of CWeaver. It provides a bidirectional communication channel between the other two components and is responsible for instantiating the intents described in the LARA aspect by either querying or changing the AST on Cetus.

This engine has a hierarchic structure that mimics the join point tree defined in the language specification. If the join point model defines that it is possible to select loops from inside a function, then there is a class, CFunction, from which it is possible to select instances of another class, CLoop. Similarly, this structure also follows the language specification regarding the join point attributes with each class implementing getter methods for specific attributes. For instance, CLoop has a *getNumIterations()* method, as the language specification defines that it is possible to consult the number of iterations of a loop join point.

The actions defined in the language specification are performed by the Weaving Engine by interacting with the AST created by Cetus. For instance, *NormalizeReturn* is a transformation pass that changes the program structure so that every function has at least one return statement. This is a transformation that exists on Cetus and was adapted to be called from a LARA aspect. Consider also *UnaryExpansion*, that transforms expressions that use unary increment or decrement operators into their equivalent binary expressions. This transformation pass was written for CWeaver as it facilitates code analysis and instrumentation.

4.3 Cetus

Cetus is a source-to-source compiler infrastructure for the C language, that supports the ANSI C standard and is written in Java. This is an infrastructure aimed towards research on multicore computing with a focus on auto-parallelization. Current work [6] tries to translate shared memory programs written in OpenMP, into other models, such as Cuda. This is supported by several transformation and analysis passes, e.g., Control Flow and Data Dependence Analysis.

The AST, the main element of Cetus regarding the CWeaver, represents the C program and is composed of Java classes such as Program, Translation Unit, Procedure, Statement and Expression. These are the basic building blocks of the C hierarchy. All these elements are used as nodes of the AST and can be visited during tree traversals. During these traversals it is possible to manipulate the nodes, effectively changing the resulting C program.

Any join point that is created during the execution is based on this AST and has a reference to the specific node it represents, e.g., a CFunction has a reference to a Procedure on the AST. Whenever there is a query about a join point attribute, a method is triggered that uses the tree reference to return the requested information. Analogously, when an action is to be performed that alters the AST, a method of the join point is called that uses the tree reference to perform the desired changes.

4.4 Join points and the AST

There is a logical match between the join points used by the LARA Engine and the AST built by Cetus. Although this is not how it is actually implemented, as there is no join point tree at any time during execution, it is the most intuitive way of thinking about the relation of both representations. For instance, there is a correspondence between a TranslationUnit and a CFile join point as both represent a C source file. With rare exceptions, mainly implementation related, there is a direct, one-to-one mapping from a join point to a node on the AST. Thus, we can think of having two trees that can be translated into one another, creating a simple, yet effective and intuitive representation of the program information.

In actuality, only the AST exists and the join points are created as needed. This is an implementation choice, taken because we want to keep the tree updated with the changes that are applied, so that any change on the tree is immediately seen by the following information queries. Consequently, it is simpler to keep a single tree, halving the updates and avoiding synchronization tasks. As expected, this choice comes with its own disadvantages, namely, having to create unnecessary join points. Picture an aspect that does not perform any change on the tree and is used only to collect information. In this case, certain join points are created repeatedly when there is no need to. Ideally, CWeaver would analyze the LARA aspect to check whether it contains tree-altering actions and, based on this information, select one of the two representations. This, however, might not be possible due to the highly dynamic nature of LARA.

5 Instrumentation Examples

This section presents two instrumentation examples that illustrate the features and advantages of using CWeaver over either extending an existing compiler or performing the tasks by hand. The first example shows a possible timer task and compares the usage of CWeaver against the usage a source-to-source compiler, while the second example presents a more complex instrumentation task that creates a call graph and is easily solved with LARA.

5.1 Loop Timer

The goal of this instrumentation is to time the execution of loops found on the code, an example taken from [6]. To this end, we use two fictional functions, *tic* and *toc*, that start and stop the timer for a loop respectively. We insert a call to *tic* before the loop and a call to *toc* after the loop. Both calls have as argument a unique identifier for the loop.

It is possible to see in Fig. 3 (a) part of the Java code that is used to implement this example in Cetus. Bear in mind that this requires changing and re-compiling Cetus' source code. While a compiler developer is more than comfortable to do so, this might not be suitable for the common user. The presented code traverses the AST, starting at the *Program*, the root node. Using a depth first iterator, it visits every node and calls *insertTimingCalls* on each *ForLoop* it finds. In order to have a complete working example we would also need to register the transformation pass in the command line options, so that it becomes accessible to the exterior.

On the other hand, if we use CWeaver, we need to write a LARA aspect that performs the code injection for us, similarly to what the Java code does. This aspect is depicted in Fig. 3 (b). Having only 9 lines of code, it produces the same output as the code shown in Fig. 3 (a). It is shorter and, to anyone with minimal experience with LARA, admittedly easier to understand. Because LARA is quite expressive and powerful when it comes to code instrumentation, the benefits of using it will certainly offset the difficulty of having to learn a new language. LARA aspects produce more concise and easier to understand code. We believe this is the best alternative for the average user, that has no previous compiler development experience.

```java
class Instrumenter {
  // ...
  public void instrumentLoops(Program p) {
    DepthFirstIterator iter = new DepthFirstIterator(p);
    int loop_number = 0;
    while ( iter.hasNext() ) {
      Object obj = iter.next();
      if ( obj instanceof ForLoop )
        insertTimingCalls((ForLoop)obj, loop_number++);
    }
  }
  private void insertTimingCalls(ForLoop loop, int number) {
    FunctionCall tic = new FunctionCall(new Identifier("tic"));
    FunctionCall toc = new FunctionCall(new Identifier("toc"));
    tic.addArgument(new IntegerLiteral(number));
    toc.addArgument(new IntegerLiteral(number));
    CompoundStatement parent =
          (CompoundStatement)loop.getParent();
    parent.addStatementBefore(loop,
          new ExpressionStatement(tic));
    parent.addStatementAfter(loop,
          new ExpressionStatement(toc));
  }
  // ...
}
```

(a) Java code for Cetus

```
aspectdef InstrumentLoops
  var counter = 0;
  select loop end
  apply
    insert before 'tic([[counter]]);';
    insert after 'toc([[counter]]);';
    counter++;
  end
end
```

(b) LARA aspect for CWeaver

Fig. 3: Codes that implement the loop timer instrumentation.

5.2 Call Graph

With this instrumentation we intend to create a call graph based on runtime information. The LARA aspect in Fig. 4 is used to inserted code inside a C source file that will output, by the end of its execution, the resulting call graph in *dot* format. The strategy will be explained with references to Fig. 4. We declare an array where each position is used to store the number of function calls of each $\langle caller, callee \rangle$ tuple. The *id* of a tuple, which is also its position in the array, is given by the *getId* method (line 6) of the *LaraObject* class. Given a tuple, this method will return its id, or create a new one if the tuple has not been seen

```
1  aspectdef CallGraph
2    var obj = new LaraObject();
3    // Increment the correct position after every function call.
4    select function.call end // <caller,callee> tuple
5    apply
6      var id = obj.getId($function.name, $call.name);
7      $call.insert after 'cg[ [[id]] ]++;';
8    end
9    // Declare the array before the first function.
10   select function end
11   apply
12     var total = obj.getTotal();
13     $function.insert before 'int cg[ [[total]] ] = {0};';
14     break;
15   end
16   // Print the array at the end of the main function.
17   select function{"main"}.return end
18   apply
19     $return.insert before 'printf("digraph call_graph {\\n");';
20     for (f in obj)
21       for (c in obj[f]) {
22         var line = 'printf(\"\\t' + f + '->' + c +
23         ' [label=\\"%d\\"];\\n\", cg[' + obj[f][c] + ']);';
24         $return.insert before '[[line]]';
25       }
26     $return.insert before 'printf("}\\n");';
27   end
28 end
```

Fig. 4: LARA aspect that inserts code to create a call graph.

before. So, with the first part of the aspect we select every $\langle caller, callee\rangle$ pair and increment the correct position of the array (line 7). After performing this operation, we have the total number of pairs and can use this information to declare the array before the first function in the source file (line 13). Finally, the last part of the LARA aspect, that starts on line 17, will add the code that prints the resulting dot graph at the end of the execution.

It is possible to see in Fig. 5 an excerpt of an example C application that was woven using the previous aspect. Line 2 of this figure contains the declaration of the array that is used while lines 5 and 16 have the increments for the $\langle print\_path, print\_path\rangle$ and $\langle main, dijkstra\rangle$ tuples respectively. At the end of the main function, just before the return statement, the call graph is printed in dot format.

6 Related Work

Several source-to-source compilers exist that are capable of dealing with C programs and perform code transformations and analyses. ROSE [12] is one such tool, with front-end support for C, C++ and Fortran. It is written in C++ and is developed in such a way that allows the creation of custom tools. Cetus [6] is another source-to-source compiler infrastructure with which there is some familiarity from previous work. It has an intermediate representation that is well suited for our goals, as it is close to the source code, allowing us to easily

```c
/* ... */
int cg[ 16 ] = {0};
void print_path(NODE * rgnNodes, int chNode) {
  if ((rgnNodes[chNode].iPrev!=9999)) {
    print_path(rgnNodes, rgnNodes[chNode].iPrev); cg[0]++;
  }
  printf(" %d", chNode);
  cg[1]++;
  fflush(stdout);
  cg[2]++;
}
/* ... */
int main(int argc, char * argv[]) {
  /* ... */
  for (((i=0), (j=(100/2))); i<100; ((i ++ ), (j ++ )))
  {
    j=(j%100);
    dijkstra(i, j);
    cg[15]++;
  }
  printf("digraph call_graph {\n");
  printf("\tprint_path->print_path [label=\"%d\"];\n", cg[0]);
  /* ... */
  printf("\tmain->dijkstra [label=\"%d\"];\n", cg[15]);
  printf("}\n");
  return 0;
}
```

Fig. 5: An excerpt of a woven C source file. The inserted code is highlighted.

inspect source-level features. Being written in Java, it provides both a portable application and a seamless integration with all LARA supporting tools.

AspectJ [9] provides a modular and general purpose AOP extension to the Java programming language. The aspects that define cross-cutting concerns in this language, are compiled into Java bytecode. Similarly, AspectC++ [13] provides an AOP extension for C++ programs. This tool compiles the weaved program to plain C++. LARA [3] is an AOP approach that is independent of both the weaving and the target language. With this language it is possible to create a customized join point model that can be used by several different languages, enabling the portability of aspects.

There are other current tools that use LARA as the language that controls the compilation flow. These weaving compilers differ not only on the target language but also on their goals. Harmonic [5] is a C source-to-source tool, that allows code insertion and provides actions for hardware-software partitioning with a focus on high-performance embedded systems. Harmonic is able to partition the input C code into tasks that are mapped into different devices such as Graphic Processing Units and Field Programmable Gate Arrays. To assist with the partitioning task, this tool is also able to estimate runtime costs that help decide where each task is going to be executed.

ReflectC [10] also targets C programs but uses LARA to guide the execution of several code optimization engines. Its back-end outputs binaries for different architectures as well as VHDL. This tool works with a low-level intermediate representation, that is closer to the target architecture, enabling access to information that might not be accessible at the source level. Hence, this tool is better suited for code optimizations.

Finally, MATISSE [2] is a MATLAB-to-C compiler framework that employs LARA aspects for code insertion and specialization actions such as variable type and size definition as well as function argument specialization. The main goal of using LARA is to aid in the translation of a highly dynamic language, MATLAB, into C source code that is well defined and suitable for high-level synthesis tools that target embedded systems.

7 Conclusions

This paper presents a source-to-source C compiler that effectively uses the best traits of AOP approaches to solve two problems related with source code instrumentation. First, the instrumentation process is semi-automated, eliminating most of the situations where human action can be error-inducing. Second, it efficiently deals with the creation of multiple source files resulting from the application of different instrumentation strategies, and that may reduce maintainability. To this end, an AOP approach allows us to maintain a single, original source file separated from the files that specify the strategies. These can be combined when needed, to create instrumented versions of the application.

As initially stated, CWeaver is a compiler developed for code instrumentation and excels at this task. It makes use of the expressiveness and high level features of LARA to simplify code instrumentation. This approach gives the user the possibility of selecting critical points in code and manipulate them to change the program or extract information. Even though instrumentation is the main purpose of this tool, we believe it is flexible enough to see different uses, adapting to several distinct tasks that may require a source-to-source compiler.

While there are no results from experimental evaluation, this tool has been successfully used in two projects that have separate goals and demand different features. The first, relies on CWeaver to implement several code instrumentation strategies so that their evaluation enables the understanding of how code invariants affect fault detection diagnosis. The other project used CWeaver as an intermediate tool to perform loop optimizations, as part of a toolchain that searches large optimization design spaces to improve performance on multicores.

7.1 Future Work

We intend to develop a mechanism that, based on the aspect code, is able to choose the most efficient internal representation. As of now, CWeaver is most effective when there are actual changes applied to the program, as opposed to when the aspect simply retrieves information, in which case, several resources are unnecessarily instantiated. As one of the next steps, we also want to perform a twofold evaluation of the tool, as no proper testing was conducted yet. Initially, we want to test the tool as a standalone compiler and measure its resource usage and compilation times, while also assessing the correctness of its output. Then, we want to evaluate how CWeaver affects the productivity of instrumentation processes. This is a non-trivial task, that needs the collaboration of several users to determine how CWeaver compares against other tools or even manual approaches.

References

1. AutoSeer: Automated Test Oracles for Software Error Detection, Please see: http://autoseer.fe.up.pt/
2. Bispo, J., Pinto, P., Nobre, R., Carvalho, T., Cardoso, J.M., Diniz, P.C.: The MATISSE MATLAB compiler. In: 11th IEEE International Conference on Industrial Informatics (INDIN). pp. 602–608. IEEE (2013)
3. Cardoso, J.M., Carvalho, T., Coutinho, J.G., Luk, W., Nobre, R., Diniz, P., Petrov, Z.: LARA: An Aspect-oriented Programming Language for Embedded Systems. In: Proceedings of the 11th Annual International Conference on Aspect-oriented Software Development. pp. 179–190. ACM (2012)
4. Cardoso, J.M., Coutinho, J.G.d.F., Carvalho, T., Diniz, P.C.: The LARA Language. In: Compilation and Synthesis for Embedded Reconfigurable Systems, pp. 35–70. Springer (2013)
5. Coutinho, J.G., Carvalho, T., Durand, S., Cardoso, J.M., Nobre, R., Diniz, P.C., Luk, W.: Experiments with the LARA Aspect-oriented Approach. In: Proceedings of the 11th Annual International Conference on Aspect-oriented Software Development Companion. pp. 27–30. ACM (2012)
6. Dave, C., Bae, H., Min, S.J., Lee, S., Eigenmann, R., Midkiff, S.: Cetus: A Source-to-Source Compiler Infrastructure for Multicores. Computer 42(12), 36–42 (2009)
7. Elrad, T., Filman, R.E., Bader, A.: Aspect-oriented Programming: Introduction. Communications of the ACM 44(10), 29–32 (2001)
8. Irwin, J., Kickzales, G., Lamping, J., Mendhekar, A., Maeda, C., Lopes, C.V., Loingtier, J.M.: Aspect-oriented Programming. In: ECOOP'97 — Object-Oriented Programming, Lecture Notes in Computer Science, vol. 1241, pp. 220–242. Springer Berlin Heidelberg (1997)
9. Kiczales, G., Hilsdale, E., Hugunin, J., Kersten, M., Palm, J., Griswold, W.G.: An Overview of AspectJ. In: ECOOP 2001 — Object-Oriented Programming, pp. 327–354. Springer (2001)
10. Nobre, R., Cardoso, J.M., Olivier, B., Nane, R., Fitzpatrick, L., Coutinho, J.G., Someren, H., Sima, V.M., Bertels, K., Diniz, P.C.: Hardware/Software Compilation. In: Cardoso, J.M.P., Diniz, P.C., Coutinho, J.G., Petrov, Z.M. (eds.) Compilation and Synthesis for Embedded Reconfigurable Systems, pp. 105–134. Springer New York (2013)
11. Nobre, R., Pinto, P., Carvalho, T., Cardoso, J.M.P., Diniz, P.C.: LARA-based Strategies for Source-to-Source Transformations Targeting Multicore Architectures. In: 17th Workshop on Compilers for Parallel Computing (CPC) (2013)
12. Schordan, M., Quinlan, D.: A Source-To-Source Architecture for User-Defined Optimizations. In: Modular Programming Languages, Lecture Notes in Computer Science, vol. 2789, pp. 214–223. Springer Berlin Heidelberg (2003)
13. Spinczyk, O., Gal, A., Schröder-Preikschat, W.: Aspectc++: An aspect-oriented extension to the c++ programming language. In: Proceedings of the Fortieth International Conference on Tools Pacific: Objects for Internet, Mobile and Embedded Applications. pp. 53–60 (2002)

A 3D cross-platform for web-based visualization of spatiotemporal information in real-time

Leonel Dias[1]

[1] INESCTEC, DEI, Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias s/n 4200-465 Porto, Portugal

pro11006@fe.up.pt

Abstract. WebGL is a new web technology that brings hardware-accelerated 3D graphics to the browser without installing additional software. A spatiotemporal database embodies spatial and temporal concepts, it captures spatial and temporal aspects of the data, and deals with the geometry and location of objects changing over time. This paper describes an exploratory work where we create an open-source framework (cross-platform) for creating a 3D virtual globe environments and interactively exploiting of spatiotemporal information in real-time using HTML5 and WebGL. For Three-Dimensional Geographic Information System (3D GIS) modeling, our hypothesis was that we would be able to generated and updated a globe 3D using techniques of computer graphic. For visualization of spatiotemporal data, our hypothesis was that we would be able to visualize large multidimensional spatiotemporal datasets with thousands of entries, each defined by a location, a time, and other attributes. To demonstrate this approach, a prototype has been developed based on the JavaScript 3D library integrated with Twitter API project. Our results indicate the proposed platform can easily support the main goals and prove the initial hypotheses.

Keywords: Computer Graphics; WebGL; Geographic Information System; Spatiotemporal; Real-time; Web-based

1 Introduction

Over the past few years the web technologies have been exploding and the capabilities for software on the web is soon to meet that of desktop software if innovation continues along this way. WebGL [1] is a JavaScript API for rendering interactive 3D graphics in modern web browsers without the use of plugins. On the other hand, ThreeJS [2] is built on top of WebGL, and allows you to create complex 3D scenes with a few lines of JavaScript. The ThreeJS supports different renders like WebGL, Canvas, SVG and CSS 3D and also supports different light sources that have specific behavior and uses.

The Three-Dimensional Geographic Information System (3D GIS) has become an important tool for interactively visualizing and investigating georeferenced content (e.g., Google Earth [15] and Esri CityEngine [16]). Emerging internet technologies, such as WebGL, offer new possibilities to develop innovative and immersive experiments in terms of geographic navigation.

The spatiotemporal databases also have become very important in recent years, as many real world applications like, for example, Geographic Information Systems (GIS), need to store real world data which shows spatial as well as temporal characteristics. A temporal GIS aims to process, manage, and analyze spatiotemporal data, as we can see in [12], where this paper describes the design and implementation of a spatiotemporal GIS database for monitoring glacier geometry and geometric change.

Thus, the main motivation of this work is to develop a cross-platform 3D for web-based visualization of spatiotemporal information in real-time. The main objectives proposed are as follows:

- Creating an interoperable, modular and highly extensible platform that facilitates the implementation of a 3D GIS Web application taking advantage of the potential offered by WebGL;
- Allow the automatic building of 3D interactive and immersive maps, with navigation and visualization functionality;
- Representing 3D georeferenced objects with different shapes and associated spatio-temporal events;
- Feeding the application dynamically and remotely, using heterogeneous information extracted in real time coming from different sources.

The paper is organized as follows: sections 2 introduces the related work; section 3 gives a full description of the platform proposed with this work; section 4 presents the experimental setup, proof of concept and discusses the results; the conclusions are presented in section 5; finally, section 6 concludes the paper indicated some future work and potential improvements.

2 Related Work

The paper [3] discusses the good and bad aspects of WebGL and introduced a way to implement a 3D WebGIS system which was based on WebGL technology. The system proposed in this paper used Ellipsoidal Mercator projection and WGS84 coordinates, made JSON file format for network transformation, described popular method for tile map service, and finished a prototype system of 3D WebGIS. As a proof of concept system was applied in a Remote Sensing Catalogue Service system, but does not implement any solution to information integration and presentation in real time.

The term real-time WebGIS can refer to different kinds of applications and parallel processing methods in Geographic Information Systems are traditionally used to accelerate the calculation of large data volumes with sophisticated spatial algorithms. So the work discussed in [4] presents a method to enable this approach for WebGIS applications. This approach uses the JavaScript 3D graphics API (WebGL) to perform client-side parallel real-time computations of 2D or 2.5D spatial raster algorithms on the graphics card. An example hillshade algorithm is described and used to demonstrate and measure the performance of the approach. Furthermore and one more time, some limitations of WebGL based spatial raster processing are discussed.

In papers [5], [6] and [7] the authors present the OpenWebGlobe project [8]. OpenWebGlobe is an open source framework for creating WebGIS 3D with a rich 3d geoinformation in web browsers using HTML5 and WebGL. In [5] and [6] The OpenWebGlobeSDK architecture is described and the authors also discuss the OpenWebGlobeSDK implementation for the creation of large scale virtual 3D globes with detailed contents and their interactive visualization directly within a broad spectrum of Web browsers. In [7] it is reported that unlike other (web-based) 3d geovisualisation technologies and toolkits, this proposed not only supports the content authoring and web visualization aspects, but also the data processing functionality for generating multi-terabyte terrain, image, map and 3d point cloud data sets in high-performance and cloud-based parallel computing environments. These papers present a good scenegraph model to WebGIS 3D.

About support the interaction of a group of users in WebGL, the authors of [9] and [10] presents two prototypes based on WebGL and WebSockets, whose chosen methodology seems to be very promising to display information in real-time.

3 Platform

News web technologies provide opportunities for new services. So, we present a new platform that allows the creation 3D maps (Globe / Earth), the representation of georeferenced objects with temporal animations, and real-time delivery and also the creation of new objects that map spatiotemporal information received during the application execution. A specific framework *_3DGIS* based on JavaScript was developed in order to implement the proposed platform. This framework is composed of several implementation methodologies that together are an asset to the design of the various services available in each Web Render Client. This framework allows to quickly implement the proposed platform in any case study, thus resulting in a generic solution totally adaptable.

3.1 Architecture of Platform

The platform's architecture consists of components that enable the implementation of services related to this work. This architecture (Figure 1) is divided into Web Render Clients and Web Server.

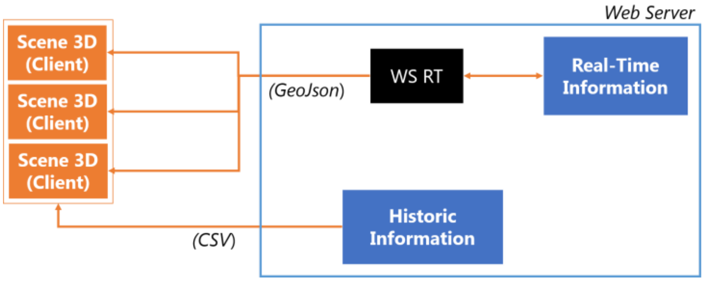


Fig. 1. The Architecture of Platform

The Web Server was developed in ASP.net MVC 4. The ASP.NET team has created lots of new goodies. One of the most useful, SignalR, is an async library for .NET to help build real-time, multi-user interactive web applications. We used this new library named ASP.NET SignalR [13] to add real-time web functionality to our platform. So, we want to inform the web client 3D when something happens (e.g., arrival of new data). For such, we write some server-side code (WS RT) that will send the message to whatever our clients may be listening to this particular event. With this methodology, our server creating a persistent connection between web server and clients that facilitate push real-time notifications.

On the other hand, the web clients was developed in a merging between WebGL, HTML5 and JavaScript through of ThreeJS. This client's application implements the management and visualization of 3D Scene, which is described in the following section.

3.2 Structure of 3D Scene

The structure of the scene is based on a scenegraph like a model and it uses an object oriented approach. The scene is the main container used to store and keep track of the objects we want to render. The renderer is responsible to render the scene in the browsers. As we can see in Figure 2, in this prototype we have one camera, two lights, one earth, various georeferenced objects and a universe.

The unique camera that we have, determines what we will see when we render the scene. We are using a perspective camera, which is what we usually associate with seeing the world. This camera has a navigation controller, which allows navigation through the scene with the mouse as input. We have added the trackball control plugin of ThreeJS which allows rotate, zoom and pan the camera.

This scene uses both ambient and directional lights. The first one to create basic light which is applied globally. The dimmed ambient light shows areas away from the "sun". And the last one to mimics the sun, at the same position as our camera.

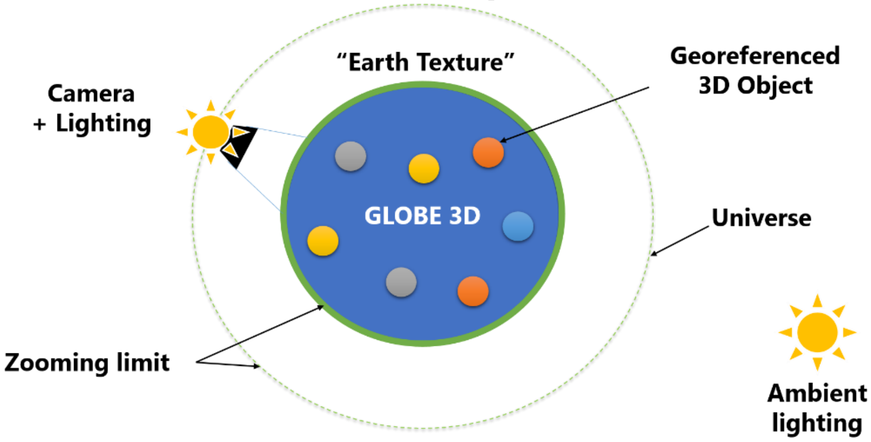


Fig. 2. Structure of 3D Scene

We start with a very basic earth. This earth is rendered as a perfect sphere (which the earth in reality isn't), where we add a texture that is a satellite map of the earth. To model earth we are using sphere geometry and material, which is referred to as a mesh in ThreeJS. The sphere is drawn as a polygon mesh and created using THREE.SphereGeometry. We use THREE.MeshPhongMaterial (Fig. 3) to wrap map data around the sphere.

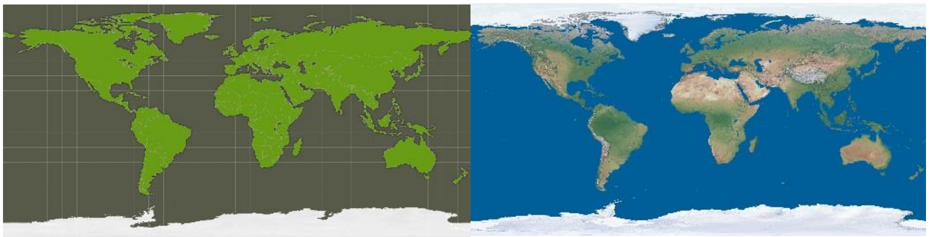


Fig. 3. Two types of layers to apply the texture of the sphere of the Earth (source [11])

This material is used to create shiny materials, and we used it to make the ocean reflective. The detail of earth in this phase is not very important, but lastly, we want to make the ocean and lakes reflective by applying a land/water mask (Fig. 4). This specular map defines the surface's shininess. Only the sea is specular because water reflects water more than earth.

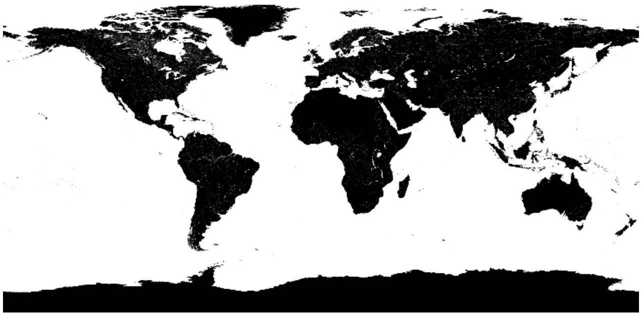


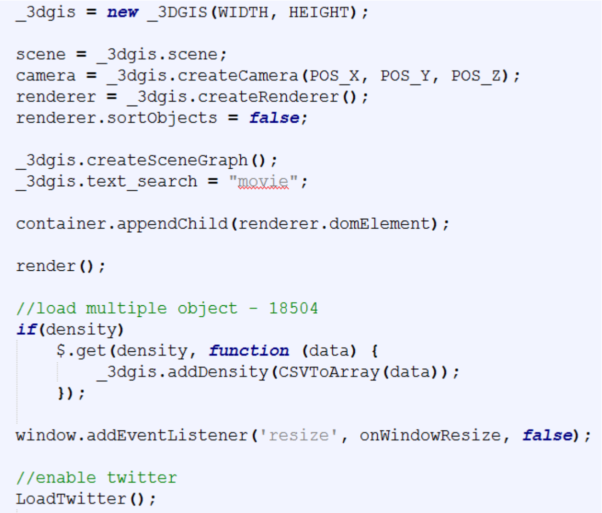
Fig. 4. Land/water mask of the Earth to create specular map (source [11])

The last thing we will add to this basic scene is a universe. The universe is created by adding a very large sphere around the Earth with the same center and project the star texture on the backside of the sphere.

With the scene rendered we need to get some information that we can use to plot on this globe. In this framework we offer possibility to load csv file, which may contain a data point with latitude/longitude (in decimal degrees format). For each coordinates pair, we need to convert this to a point on our earth. So, we implemented a function that converts a coordinates of a point in a 3D space. This, function receives point's coordinates, radius of our earth and the height is used as an offset of how high above the surface we want to start drawing. With this method, after reading a simple preconfigured csv, can put and render anywhere from our Earth any georeferenced objects 3D. For each object with spatiotemporal attributes associated, we can provide temporal animations to highlight objects recently created and associated to events that have emerged from information collected in real time. In this work, we use tween.js from ThreeJS and we allow use it to easily smooth your animation. This engine provides tweening and full blown animations thanks to chaining.

4 Results

As a proof of concept, we have developed a prototype platform named *"GIS 3D Real-Time"*. To be able to display something with framework _3DGIS developed, the following code (Figure 5) is the basic code we need to get started.



```
_3dgis = new _3DGIS(WIDTH, HEIGHT);

scene = _3dgis.scene;
camera = _3dgis.createCamera(POS_X, POS_Y, POS_Z);
renderer = _3dgis.createRenderer();
renderer.sortObjects = false;

_3dgis.createSceneGraph();
_3dgis.text_search = "movie";

container.appendChild(renderer.domElement);

render();

//load multiple object - 18504
if(density)
    $.get(density, function (data) {
        _3dgis.addDensity(CSVToArray(data));
    });

window.addEventListener('resize', onWindowResize, false);

//enable twitter
LoadTwitter();
```

Fig. 5. An example implementation using the developed framework (*_3DGIS*)

With previous code, we can display a prototype like shown in figures 6 and 7. In this example, the framework developed was tested within two different georreferenced dataset. The visualization platform is designed for a dynamic and intuitive usage, aimed at revealing the georeferenced information in an understandable and interactive manner. In Figure 6 we can see a feature of the prototype implemented, which has hundreds of small cubes, wherein each is a georeferenced information stored in a CSV file.

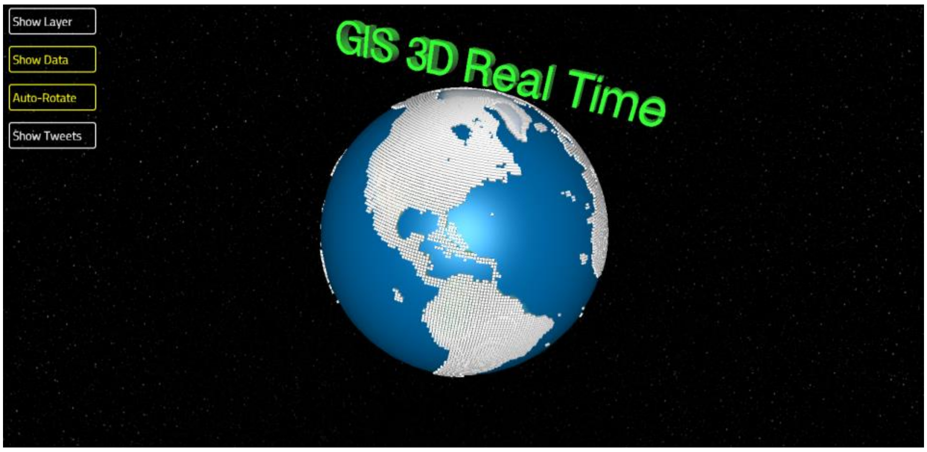
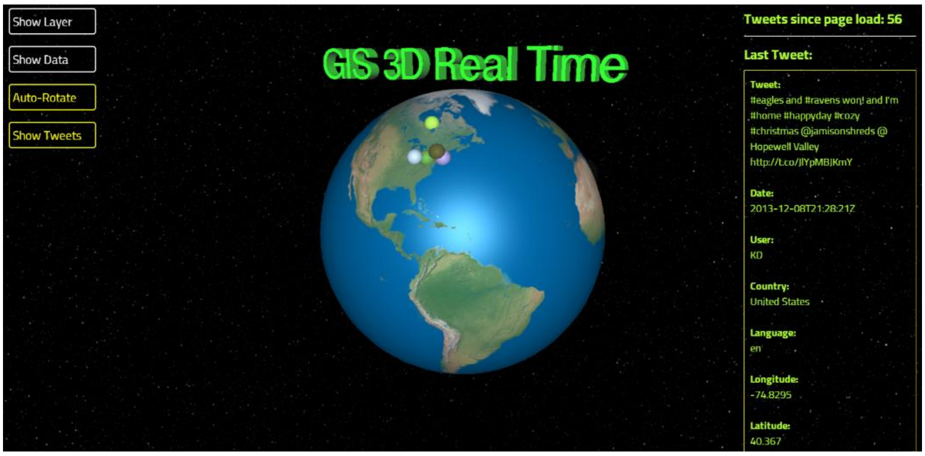


Fig. 6. Proof of Concept (Load CSV File)

In Figure 7 we can see another feature of the prototype implemented, which displays in real time the tweets that are posted by users of Twitter. This case study accesses the Twitter Streaming API [14] to collect and display real time tweets. In this last case, there is spatiotemporal information therefore we can see temporal animations associated with each object, it represents a tweet. As seen in the same figure, each tweet has a simple template and presentation itself with features ranging from the enumeration of its key features, according to parameters select in the framework.



Tweets since page load: 56

Last Tweet:

Tweet:
#eagles and #ravens won! and I'm #home #happyday #cozy #christmas @jamisonshreds @ Hopewell Valley http://t.co/JlYpMBJKmY

Date:
2013-12-08T21:28:21Z

User:
KD

Country:
United States

Language:
en

Longitude:
-74.8295

Latitude:
40.367

Fig. 7. Proof of Concept (Load Tweets in Real-Time)

As seen in figures 6 and 7, at this moment, the first interface that the user has access is the main menu on the top left of scene view. From this menu, it has access to four submenus: "Show Layer", "Show Data", "Auto-Rotate" and "Show Tweets". The menu

"Show Layer" allows him to toggle the map layer (“texture of earth”) to show different data overlays (for instance, switching between the two layers of figure 3).

5 Conclusions

At this point, it is premature to present an overall evaluation of the platform. Since its practical implementation has not yet been completed. This platform constitutes work still under progress, yet it already provides very interesting and powerful features. However, after the work done, it is considered that the proposed platform addresses the initial goals.

We must note the fact that some of the features implemented are still in a testing phase and is therefore necessary to develop testing standards covering the evaluation of the solution in terms of performance, sustainability load and interaction with users.

The main problems present in this platform are: some parameterizations are still hardcoded and do not allow great customization to the users; Too much time spent to on loading heavy textures (layers) of the Earth, which lowers the rapidity startup of the platform; Each web client application does not notify the web server (that distributes information) which left to wait updates (may lead to the server overload).

6 Future Work

Many features are still to come, and this platform is yet to achieve its full potential with this approach. Future work comprehends finishing and integrating this platform with other features of the Web GIS 3D. E.g., it is expected that in the near future this platform become an aggregator of information may eventually be georeferenced provided by different sources within the geospatial search engine. The new features in the future can also offer the user possibility to fly to any place around the earth and see 3D buildings, imagery and terrain, like as Google Earth [15].

Another future development may be taking advantage of the emerging browser functionalities and mobile devices. These future developments will follow the current trends of mobile access, thereby enabling to extend the platform to be more easily used in mobile devices. E.g., build extension mobile platform taking advantage of Global Positioning System (GPS) sensors integrated in some smartphones.

REFERENCES

1. OpenGL ES 2.0 for the Web. Last visited 28 December 2013. http://www.khronos.org;
2. ThreeJS - JavaScript 3D library. Last visited 5 December 2013. http://threejs.org;
3. Lei Feng; Chaoliang Wang; Chuanrong Li; Ziyang Li (2011). "A Research for 3D WebGIS based on WebGL," Computer Science and Network Technology (ICCSNT), 2011 International Conference on , vol.1, no., pp.348,351, 24-26 Dec. 2011;
4. Auer, M. (2012). Real-Time Web GIS Analysis Using WebGL. International Journal of 3-D Information Modeling (IJ3DIM), 1(3), 49-61;

5. Christen, M., Nebiker, S., & Loesch, B. (2012). Web-Based Large-Scale 3D-Geovisualisation Using WebGL: The OpenWebGlobe Project. International Journal of 3-D Information Modeling (IJ3DIM), 1(3), 16-25;
6. Christen M. & Nebiker S. (2011), OpenWebGlobe SDK, an Open Source High Performance Virtual Globe SDK for Open Maps. In Manuela Schmidt and Georg Gartner, editors, Proceedings of the 1st European State of the Map;
7. Loesch B., Christen, M. and Nebiker, S., Flückiger K. (2012). OpenWebGlobe – An Open Source SDK for Creating Large-Scale Virtual Globes on a WebGL Basis, ISPRS 2012, Melbourne;
8. OpenWebGlobe | WebGL geospatial data viewer. Last visited 28 December 2013. http://www.openwebglobe.org;
9. Bijin Chen; Zhiqi Xu, (2011) "A framework for browser-based Multiplayer Online Games using WebGL and WebSocket," Multimedia Technology (ICMT), 2011 International Conference on , vol., no., pp.471,474, 26-28 July 2011;
10. Charles Marion and Julien Jomier (2012). Real-time collaborative scientific WebGL visualization with WebSocket. In Proceedings of the 17th International Conference on 3D Web Technology (Web3D '12). ACM, New York, NY, USA, 47-50;
11. Planet Earth Texture Maps. Last visited 5 December 2013. www.planetpixelemporium.com.
12. Mennis, J. L. and Fountain, A. G. (2001). A spatio-temporal GIS database for monitoring alpine glacier change. Photogrammetric Engineering and Remote Sensing, 67: 967–975;
13. ASP.NET SignalR | Incredibly simple real-time web for .NET. Last visited 27 December 2013. http://signalr.net;
14. Twitter. Last visited 28 December 2013. http://www.twitter.com;
15. Google Earth. Last visited 28 December 2013. http://www.google.com/earth;
16. Esri CityEngine | 3D Modeling Software for Urban Environments. Last visited 28 December 2013. http://www.esri.com/software/cityengine;

Dungeon multi-level generator for a serious game

Ricardo Gonçalves, António Coelho

INESCTEC, DEI, Faculdade de Engenharia, Universidade do Porto,
Rua Dr. Roberto Frias s/n 4200-465, Porto, Portugal

refg@fe.up.pt; acoelho@fe.up.pt

Abstract. Game content creation tends to be an expensive time-consuming task, however by using procedural generation algorithms, it is quite possible to quickly generate content that is of good enough quality and in some cases can even offer a greater replay-ability. By integrating these benefits into a serious game for learning computer programming, which is currently under development, we hope to create a versatile tool that will help students. To this end several algorithms were evaluated and one was adapted to the needs of the game. As a result, the implemented algorithm was able to generate interesting multi-level dungeons that are adequate to the serious game it will integrate.

Keywords. Procedural Content Generation in Games; Serious Games; Dungeon Crawler; Level Generation

1 Introduction

Serious games are slowly becoming an alternative method for imparting knowledge and skills, both in academic and enterprise environments, although they will never totally replace formal education, but act more as a complement. This technology, as the name implies, uses game characteristics such as rules; challenges; rewards; interaction and feedback, and wraps these around a serious concept. By achieving a balance between fun and the serious subject, these types of games can become motivational force to engage learners to learn the desired knowledge and skills and if possible go beyond the initial scope.

In spite of the intrinsical and extrinsical motivational factors inherent to any game, whether serious or not, one of the biggest engagement pit falls has always been their longevity and replay-ability. Once the novelty wears off, players rarely feel the need to come back, especially if it is a single player game. This is where the Procedural Content Generation in Games (PCG-G) may help, by generating content in real-time, either as the game evolves or every time a new game is initiated.

In addition to the replay-ability potential that PCG-G algorithms may introduce, there is also a great advantage of reducing the workload on the design teams by either totally generating part of the content or generate it to a point where artists can pick the generated assets and polish and add further detail, enriching the overall quality of the application and reducing the time and cost it would take if content was done manually from scratch [1].

This paper explores the initial development of an algorithm that generates dungeon like levels for a serious game for the teaching of computer programming [2]. At the moment maps are handmade which makes it a huge time-consuming task and offers little to no replay-ability, plus every student gets the same scenario.

This paper is organized as follows. Section 2 defines the context of PCG-G used and presents some base algorithms for dungeon generation. In section 3 the algorithm to be used is specified thoroughly and finally, the conclusions and future work are presented.

2 State of the Art

This section will begin to explain what procedural content generation in games is and what type of content can be generated as well as the distinction between adaptive and non-adaptive PCG algorithms. Finally, a few PCG algorithms will be explored.

2.1 Procedural Content Generation in Games

Procedural Content Generation in games (PCG-G) can be defined as the automatic creation of content, which is achieved through the implementation of an algorithm tailored to a specific end [3]. This process can be achieved through a random process or through deterministic reconstruction based on parameters which will create the same object every time. Since the main focus of this article is the creation of continuous new experiences in game environments, the focused PCG-G algorithms will have some degree of stochasticity. Note that the random number generator seed should not be considered an algorithm parameter, otherwise every algorithm would have to be considered deterministic.

Even though content is intended to be greatly different in every generation iteration, it is a good policy not to be completely random. Instead PCG-G algorithms should implement some degree of stochasticity that adheres to constraints, which can be affected by associated parameters, of what content can be generated and how [4]. By generating content randomly within that range of constraints, the result will be adequate and within the scope of the game. Furthermore, by allowing the player to act, directly or indirectly, upon the parameterization of the algorithm the results will be tailored to the choices he made [5]. These algorithms fall under the category of adaptive-PCG and can be useful, for instance, to adjust the difficulty of a generated level to the player's proficiency.

But what type of game content can be generated by PCG algorithms? Theoretically, almost anything can be generated content. Hendrikx *et al.* [1] defines four classes of game content that can be generated procedurally:

- **Game Bits** include textures; item properties; sounds; vegetation; buildings; behavior; weather and finally natural elements.

- **Game Space** pertains to the environment present in-game, whether they are indoor (dungeons; rooms; houses; etc.) or out-door (forests; space; underwater; etc.).
- **Game Systems** such as ecosystems; road networking; urban environment development and entity behavior management.
- **Game Scenarios** define puzzles; story and level concept.

Another important aspect to take into consideration is whether PCG-G algorithms are run online or offline [3]. A PCG-G online algorithm is usually executed in runtime, as the game progresses, generating new content on the fly. For example Minecraft[1] is continuously generating the environment as the player moves into uncharted territory. Online algorithms may be adaptive and have two critical requirements. The first of these requirements is speed of execution because if these algorithms take too long doing the necessary computations it will have a detrimental effect on the player's immersion. As for the second requirement, the algorithm needs to assure that a minimum level of quality is attained.

On the other hand, offline algorithms are usually used during development and subject manual polish to meet the desired results and as such, these are almost always non-adaptive.

2.2 PCG-G Algorithms for Dungeon Generation

Game level generation usually have specific constraints that are associated with the type of level being generated which usually requires a specific oriented algorithm. In this sense an algorithm for generating an open world with oceans, mountains, etc. may not be the most adequate to generate a dungeon area. For this reason, in this section, some base algorithms will be presented which are mainly oriented to generate dungeon levels.

2.2.1 Random Room Placement

This algorithm consists mainly in populating a designated gridded area with rooms of random size and connecting them through corridors. As parameters this algorithm can receive the number of rooms to generate, width and height ratio. The general work flow of this algorithm is as follows: firstly it starts by parsing the parameters, proceeding to a loop state where rooms are generated. For each room generation loop iteration, the generated room must be first validated before placing it on the map. For example if the room overlaps another or is outside the bounds of the map it is discarded and the room is redone. This process loops until all rooms have been generated or until some other exit condition is met, such as reaching the maximum number of attempts. After all rooms have been successfully generated the next step is the creation of the corridors. This step usually involves using a path finding algorithm such as A* to find the closest path from room to room. When all rooms have been connected, if

[1] https://minecraft.net/

there are still unconnected doors in some of the rooms, additional dead end corridors may be added.

2.2.2 Space Partitioning For Room Placement

In the previous example, during room generation and map placement, it could occur with some frequency that rooms would get discarded. With space partitioning algorithms [6], such as BSP trees and quadtrees this problem is circumvented because rooms are placed in the empty generated areas and therefore do not overlap other rooms.

BSP tree random placement starts by sub dividing the initial map area into smaller sections by choosing a horizontal and vertical direction. This sub-division is done for n iterations. After reaching the n iterations rooms are inserted into the subdivided areas. Note that rooms must be of the same size or smaller than the regions they will be placed on. After all rooms have been successfully placed, the next step is connecting the rooms by looping through all split regions and connecting each immediate neighboring regions ensuring that all rooms get connected.

Similarly to BSP trees, quadtrees start to divide a large area but in this case it divides it in four equal squares. For each new area it is again subdivided until either an area reaches the minimum room size or if the stochastic element of the algorithm stops subdivision. For each leaf a room is placed such that it is entirely contained within the leaf area.

2.2.3 Cellular Automaton Method

While the two previous algorithms were more oriented to generate room-like dungeons, a cellular automaton algorithm is more oriented to generate cavern-like areas [7]. The first step in this algorithm is to fill an area randomly with walls and empty spaces. Subsequently the algorithm parses each of the grid cells and applies the *4-5 rule*: a cell C at position p becomes a wall if at least five (including itself) plus the eight immediate connected neighboring cells are walls. Parsing of each the cells need to be done simultaneously instead of parsing one by one and using the value of the previous to calculate the next. The reason for this is purely due to the fact that it will look more organic and less artificial. One of the problems this algorithm has is that it tends to have very inconsistent results such as wide open areas or disjoint areas. This problem can be solved however by adding an additional refinement rule before applying the initially defined rule[2]. This new rule contains the first one plus: a cell C becomes a wall if two or less cells within 2 steps from C are walls. A mathematical notation of these rules can be for example:

Rule 1: $W'(p) = C_1(p) \geq 5 \; or \; C_2(p) \leq 2$

Rule 2: $W'(p) = C_1(p) \geq 5$

[2] http://pixelenvy.ca/wa/ca_cave.html

Where $W'(p)$ represents if there is a wall in a particular position p, and $C_n(p)$ represents the number of cells within n steps of the position p are walls.

2.2.4 Search-Based Procedural Generation

Search-based procedural generation algorithms use evolutionary biology based approaches, similar to genetic algorithms [8], where certain aspects, such as survival of the fittest and genetic recombination or mutation, are important parts of the overall algorithm.

To represent a problem using this method, individuals are defined as chromosomes, which can be subjected to either or both *crossover* and *mutation* operations. The crossover operation is a process that involves two or more parent chromosomes where specific aspects of the parents are selected and combined to form a child. This potentially allows the exploration and exploitation of the map space. The mutation process only operates over a single chromosome and changes/mutates some specific aspects.

An important aspect is how these chromosomes are selected to be operated. This selection usually involves a fitness score for each chromosome which translates into its ability to solve the problem. Usually chromosomes are subjected to these operations for several generations until specific exit conditions are met, such as attaining an overall solution fitness score.

3 Methodology

The developed algorithm which is the main focus of this paper is an online algorithm because each new game, for each player, will have different unique levels and it is therefore a requirement that maps are generated at game run-time. This algorithm is based on a random room placement approach with some differences that have been influenced by the map generation process used in the game TinyKeep[3]. The reason behind this choice was strongly due to the aesthetic look and topology that could be achieved with this algorithm, which feels more varied and less artificial, consequently having more immersion potential which is one of the main goals of the serious game this algorithm [9] is meant to integrate.

The first important aspect of this algorithm is the entry parameters. These parameters are the number of rooms to be generated, a set of values representing the mean and deviation parameters for a Gaussian distribution function, a max room size ratio, the minimum room size and finally the minimum boundary between rooms. These parameters will be further explained as the steps involved in this algorithm are detailed.

The algorithm is divided in different sequential steps that use the data from the previous step to improve it to feed the next until the final result is reached. The first step is randomly creating rooms and positioning them over the map area without wor-

[3] http://tinykeep.com/

rying about overlapped rooms. The room creation is basically a loop that generates as many room as set in the entry parameters. For each room generated, before accepting it, it is verified if it meets certain constraints such as room ratio, to avoid very narrow rooms for example, or if the room has a minimum size, for example to avoid single cell rooms. Rooms random generation is done by using a Gaussian distribution function to ensure that most rooms are within the same size/ration range and only a few are bigger, creating a more natural distribution and avoiding having a map filled with only huge rooms.

The next step on this process is to ensure that no rooms are overlapped and to achieve this, a simple separation steering process is used on each room until no room overlaps. This steering process uses the entry parameter minimum border, which can be equal or higher than zero, to increase (or use the default room size if it's zero) to check if two rooms overlap. The main reason behind this parameter is to prevent rooms to be closely packed, minimizing space for later corridor generation. Also, the separation steering process ensures that the room is snapped to the grid to ease the corridor generation. Because this is a multi-leveled map, or in other words, it can have infinite levels, when a new level is added to a map generated in a previously iteration of this algorithm process, there will be some special rooms called transition rooms. These transition rooms are rooms which were selected by their area (the higher the area the more chances it has to be selected as a transitional room) and they are meant to establish a transition point between the one level and the next room. In figure 1, these rooms can be seen as blue if they are transition rooms to the next room, or pink if they transition to the previous room. Upon creating a new map, the previous map transition rooms are passed and integrated within the selected rooms of the new level. These rooms remain fixed during the process of separation steering of the new level to ensure they retain their relative position to the parent's level.

After all rooms have been correctly positioned the next step is selecting the rooms that will be part of the final map level. This selection is random but also based on their areas, where bigger rooms get some higher preference. After this selection is complete, at least two rooms are selected as transition points.

With a list of selected rooms the next step is the creation of a graph that connects rooms together. This is achieved by constructing a graph similar to the a relative neighborhood graph, except that this one was done in such a way that in some cases a map could have disconnected sub graphs which could be a potentially way to allow some kind of teleportation mechanic in the game. This can be disabled by simply selecting the two closes nodes of each closest sub-graph and connecting them ensuring this way that every room is reachable by normal means. One of the side effects of this approach was that some rooms were connected by several nodes, creating multiple paths between nodes and to solve this, a minimal spanning tree is generated from the initial graphs, reducing the number of connections between nodes. This will ensure that there are no loops within the same level, but can be by using a different level.

The final step is connecting rooms using the graph. To achieve this, a path finding algorithm was used, more specifically A-star, which is used to create a route for each graph edge.

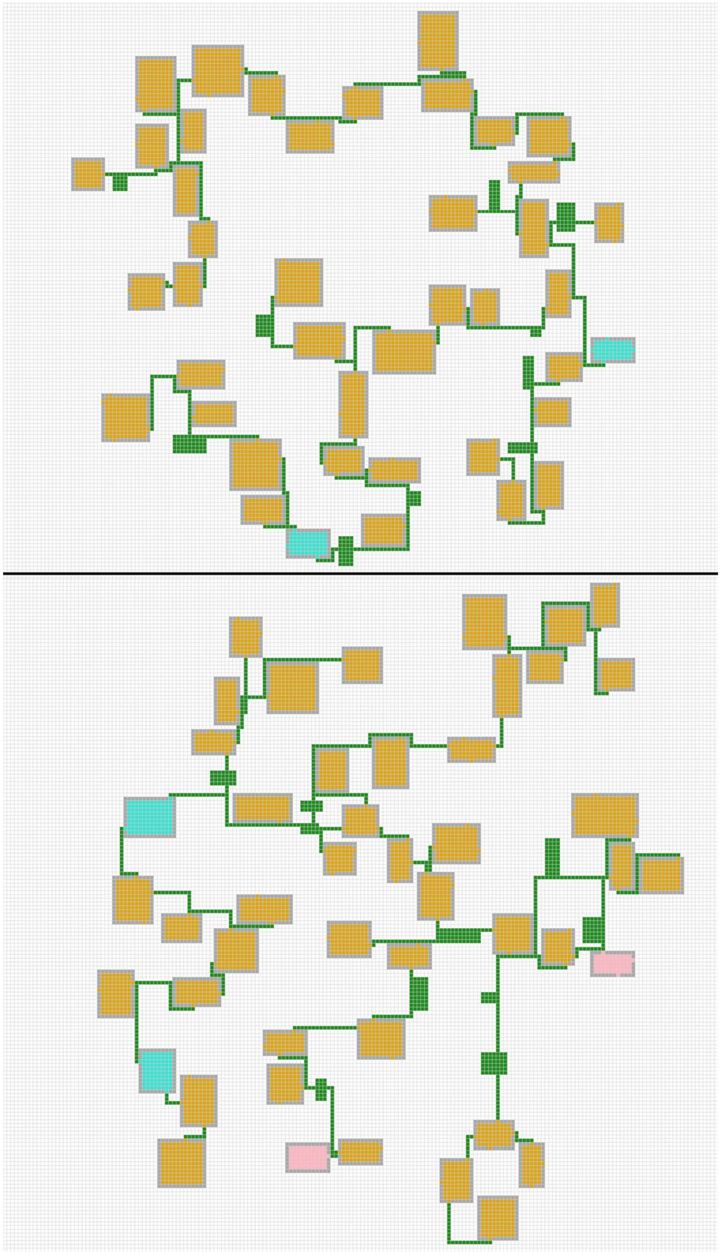


Figure 1. Example of a generated map with two levels. Blue rooms are transition rooms to the next level and pink rooms are transition rooms to the previous level.

To avoid zig-zag corridors the algorithm punishes heavily direction alteration, ensuring that corridors have a straighter format. Also the different cost values are given to cells to ensure that the algorithm will always favor existing corridors, adding more realism and variety to the map, like intersections. Finally, after a path as been found between two rooms, this path is checked against all non-selected rooms and if an intersection occurs, the room is integrated as part of the corridor, adding further detail to the general map layout.

4 Conclusions and Future work

Aesthetical and functional wise the generated maps achieve the goals. They lay the foundation for a rich dungeon environment and they ensure that there are several pathways that can be used to establish a sequence of programming exercises with different difficulty levels (ways), while also allowing in parallel to have a narrative with different aspects that can be explored.

The algorithm is not finished however, for the moments all rooms are squared which may suit the more general cases, but to further increase immersion some rooms could be improved by using pre-defined model that add more detail to the map. Another evolution is to have a pre-determined pathway that is passed to the algorithm and generate around it a map level. This could be useful in the eventuality of wanting to have a set of rooms with specific exercises and narrative plots defined by the teacher.

REFERENCES

1. Hendrikx, Mark, et al. "Procedural content generation for games: a survey." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 9.1 (2013): 1.
2. Coelho, António; Kato, Enrique; Xavier, João; Gonçalves, Ricardo: Serious Game for Introductory Programming. In Proceedings of the Second International Conference on Serious Games Development and Applications (SGDA 2011), Lisbon
3. Togelius, Julian, et al. "Search-based procedural content generation: A taxonomy and survey." *Computational Intelligence and AI in Games, IEEE Transactions on* 3.3 (2011): 172-186.
4. Togelius, Julian, et al. "What is procedural content generation?: Mario on the borderline." *Proceedings of the 2nd International Workshop on Procedural Content Generation in Games*. ACM, 2011.
5. Yannakakis, Georgios N., and Julian Togelius. "Experience-driven procedural content generation." *Affective Computing, IEEE Transactions on* 2.3 (2011): 147-161.
6. Togelius, Julian; Shaker, Noor; Nelson, Mark J. "Procedural Content Generation in Games: A Textbook and an Overview of Current Research", 2014
7. Johnson, Lawrence; Georgios, Yannakakis; Togelius, Julian. "Cellular automata for real-time generation of infinite cave levels." *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*. ACM, 2010.
8. Valtchanov, Valtchan; Brown, Joseph. "Evolving dungeon crawler levels with relative placement." *Proceedings of the Fifth International C* Conference on Computer Science and Software Engineering*. ACM, 2012.
9. Doull, Andrew. "The death of the level designer." Internet: http://pcg. wikidot. com/the-death-ofthe-level-designer (2008).

PAPERS IN ALPHABETICAL ORDER

AUTHORS IN ALPHABETICAL ORDER

DSIE Doctoral Symposium in informatics Engineering | 14

The 9th Doctoral Symposium in Informatics Engineering (DSIE'14) continues the series of past workshops at Faculty of Engineering of the University of Porto (FEUP) and will take place at FEUP on the 30th and 31th of January 2014.

The Doctoral Symposium occurs as part of the "Methodologies for Scientific Research" class of FEUP's Doctoral Program (ProDEI) with the purpose of providing students with new competences and abilities in the experience of scientific writing and assessment.

The main goal of DSIE is to provide graduate students a setting where they can informally expose and discuss their work, while collecting valuable expert opinions. The sharing of new ideas, methods and applications is also relevant and encouraged, especially because it may improve the buildup of their theses content.

-This edition of the Symposium included the following major topics:

- Virtual and Simulation Systems
- Communication Networks
- Security Systems
- High-performance Computing
- Information Systems
- Software Engineering
- Machine Learning and Intelligent System