Open University of Catalonia

Doctoral Thesis Report

An Information Security Model based on Trustworthiness for Enhancing Security in On-line Collaborative Learning

PhD candidate:
Jorge Miguel

Supervisors:
Dr. Santi Caballé
Dr. Fatos Xhafa

A thesis submitted in fulfilment of the requirements
for the degree of Doctoral Programme on Network and Information Technologies

in the

IT, Multimedia and Telecommunications Department

18th May 2015

Declaration of Authorship

I, Jorge Miguel, declare that this thesis titled, 'An Information Security Model based on Trustworthiness for Enhancing Security in On-line Collaborative Learning' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a doctoral degree at the Open University of Catalonia (UOC).
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the UOC or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 18th May 2015

"Security is both a feeling and a reality. And they're not the same."

Bruce Schneier

Acknowledgements

I would like to express my gratitude to my doctoral thesis supervisors, Dr. Santi Caballé and Dr. Fatos Xhafa, for their support, guidance, and encouragement.

I would also like to thank Dr. Josep Prieto for his assistance and support as well as the lecturers and students (Open University of Catalonia) who eagerly participated in the experiences reported in this thesis. Special thanks to the RDlab team (Technical University of Catalonia - BarcelonaTech) for the technological assistance.

Finally, and most importantly, I would like to thank my wife Iris for her unwavering support.

OPEN UNIVERSITY OF CATALONIA

Abstract

IT, Multimedia and Telecommunications Department

Doctoral Programme on Network and Information Technologies

An Information Security Model based on Trustworthiness for Enhancing Security in On-line Collaborative Learning

by Jorge Miguel

This thesis' main goal is to incorporate information security properties and services into on-line collaborative learning by a functional approach based on trustworthiness assessment and prediction. As a result, this thesis aims at designing an innovative security solution, based on methodological approaches, to provide e-Learning designers and managers with guidelines for incorporating security into on-line collaborative learning. These guidelines include all processes involved in e-Learning design and management, such as security analysis, learning activities design, detection of anomalous actions, trustworthiness data processing, and so on.

The subject of this research is conducted by multidisciplinary and related research topics. The most significant ones are on-line collaborative learning, information security, Learning Management Systems (LMS), and trustworthiness assessment and prediction models. In this scope, the problem of securing collaborative on-line learning activities is tackled by a hybrid model based on functional and technological solutions, namely, trustworthiness modelling and information security technologies.

Up to now, the problem of securing collaborative activities in distance education against unfair and dishonest on-line assessment (e.g. cheating) has been mainly tackled with technological security solutions. Over the last years, security solutions for e-Learning have evolved from specific and isolated security properties, such as privacy, to holistic models based on technological security comprehensive solutions, such as Public Key Infrastructures and multidisciplinary approaches from different research areas. Current technological security solutions are feasible in many e-Learning scenarios but on-line collaborative learning involves specific components, such as on-line assessment activities, that usually bear specific issues and challenges that e-Learning designers have to face when they manage security requirements. In this context, even the most advanced

and comprehensive technological security solutions cannot cope with the whole scope of on-line collaborative learning vulnerabilities. Therefore, to overcome these deficiencies, the solution proposed in this thesis endows security technological techniques with a functional approach based on trustworthiness in order to enhance security in on-line collaborative learning activities. Trustworthiness is closely related to interactions between agents, thus trustworthiness approaches are suitable for modelling and measuring collaborative learning interactions, in terms of trustworthiness factors, rules and features.

The research methodological approach of this thesis involves building, experimenting and validating on a comprehensive security methodology offering a guideline for the design and management of collaborative activities based on security properties and trustworthiness approaches. Security properties are analysed in the context of collaborative activities by considering specific security requirements for these activities. From this model, secure on-line collaborative activities based on trustworthiness approaches were designed in this thesis in terms of learning components in learning management systems. The design proposed enables e-Learning tutors to discover anomalous students' behaviour that compromises security in on-line learning activities by trustworthiness assessment and prediction. The detection model is supported by specific methods and techniques, such as peer-to-peer visualization tools as well as prediction based on neural networks. Finally, with the aim to evaluate the design process of on-line collaborative activities based on trustworthiness approaches, a solid plan of pilot experiments was developed in our real context of e-Learning of the Open University of Catalonia which validates the trustworthiness assessment and prediction methodology proposed in this thesis. Based on the results achieved in this thesis, future directions of research are proposed.

List of Research Contributions

The relevance of the research in this thesis is supported by 15 research contributions to several journals and international conferences. In all of these contributions, the doctoral candidate is the first and corresponding author. This section shows the main contributions. The publications provided, both journals and conference papers, were published or accepted for publication within the last four years. Finally, this research will be reported in a new author book edited by Elsevier.

The three most relevant contributions of the thesis, indexed in ISI-JCR and meeting the academic regulations and requirements, are the following journal papers:

- **Miguel, J.**, Caballé, S., Xhafa, F., and Prieto, J. (2015a). Security in online web learning assessment. providing an effective trustworthiness approach to support e-learning teams. *World Wide Web Journal (WWWJ)*. Springer. doi:10.1007/s11280-014-0320-2. IF: 1.623, Q1: 20/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE)

- **Miguel, J.**, Caballé, S., Xhafa, F., and Prieto, J. (2015b). A massive data processing approach for effective trustworthiness in online learning groups. *Concurrency and Computation: Practice and Experience (CCPE)*, 27(8):1988–2003. Wiley Online Library. doi:10.1002/cpe.3396. IF: 0.784, Q2: 50/102, Category: COMPUTER SCIENCE, THEORY & METHODS (JCR-2013 SE)

- **Miguel, J.**, Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2015c). A methodological approach for trustworthiness assessment and prediction in mobile online collaborative learning. *Computer Standards & Interfaces (CSI)*. Springer. doi:10.1016/j.csi.2015.04.008. IF: 1.177, Q2: 38/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE)

Contributions published in conference proceedings by IEEE Computer Society:

- **Miguel, J.**, Caballé, S., and Xhafa, F. (2015d). Methods and issues in visualization of trustworthy data from eLearning systems. In *Fifth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional approaches (ALICE 2015)*, Taipei, Taiwan. IEEE Computer Society. submitted

- **Miguel, J.**, Caballé, S., Xhafa, F., and Snasel, V. (2015). A data visualization approach for trustworthiness in social networks for on-line learning. In *29th IEEE International Conference on Advanced Information Networking and Applications*

(AINA-2015), pages 490–497, Gwangiu, South Korea. IEEE Computer Society. 10.1109/AINA.2015.226

- **Miguel, J.**, Caballé, S., Xhafa, F., and Prieto, J. (2014a). Security in online assessments: Towards an effective trustworthiness approach to support e-learning teams. In *28th International Conference on Advanced Information Networking and Applications (AINA 2014)*, pages 123–130, Victoria, Canada. IEEE Computer Society. 10.1109/AINA.2014.106. **Best Paper Award of AINA 2014**

- **Miguel, J.**, Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014b). A collective intelligence approach for building student's trustworthiness profile in online learning. In *2014 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2014)*, pages 46–53, Guangzhou, P.R. China. IEEE Computer Society. 10.1109/3PGCIC.2014.132

- **Miguel, J.**, Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014c). A methodological approach to modelling trustworthiness in online collaborative learning. In *Fourth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2014)*, pages 451–456, Salerno, Italy. IEEE Computer Society. 10.1109/INCoS.2014.18

- **Miguel, J.**, Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014d). Predicting trustworthiness behavior to enhance security in on-line assessment. In *2014 6th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2014)*, pages 342–349, Salerno, Italy. IEEE Computer Society. 10.1109/INCoS.2014.19

- **Miguel, J.**, Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014e). Towards a normalized trustworthiness approach to enhance security in on-line assessment. In *Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2014)*, pages 147–154, Birmingham, UK. IEEE Computer Society. 10.1109/CISIS.2014.22

- **Miguel, J.**, Caballé, S., and Prieto, J. (2013a). Information security in support for mobile collaborative learning. In *The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013)*, pages 379–384, Taichung, Taiwan. IEEE Computer Society. 10.1109/CISIS.2013.69

- **Miguel, J.**, Caballé, S., and Prieto, J. (2013b). Providing information security to MOOC: Towards effective student authentication. In *5-th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2013)*, pages 289 – 292, Xian, China. IEEE Computer Society. 10.1109/INCoS.2013.52

- **Miguel, J.**, Caballé, S., and Prieto, J. (2012b). Providing security to computer-supported collaborative learning systems: An overview. In *Fourth IEEE International Conference on Intelligent Networking and Collaborative Systems (INCOS 2012)*, pages 97–104, Bucharest, Romania. IEEE Computer Society. 10.1109/iNCoS.2012.60

In addition, we include a paper published in a relevant on-line journal as well as a book of a prestigious editorial where this thesis work will be reported:

- **Miguel, J.**, Caballé, S., and Prieto, J. (2012a). Security in learning management systems: Designing collaborative learning activities in secure information systems. *eLearning Papers. European Comission*, 28:1–3. elearningeuropa.info. ISSN: 1887-1542

- **Miguel, J.**, Caballé, S., and Xhafa, F. (2015e). *Intelligent Data Analysis for e-Learning: Trustworthiness for Enhancing Security in on-line Learning Teams and Networks*. Intelligent Data-Centric Systems. Elsevier, Amsterdam, Netherlands. Book proposal submitted

Finally, other contributions of the candidate from previous research projects.

These projects helped the candidate to improve his skills as a researcher in a great deal, but these projects were either out of the thesis scope or carried out before starting the thesis. For this reason, the contributions coming from these projects are not listed in this section though they were included in the Bibliography section: Miguel (2005a,b, 2006a,b, 2007a,b,c, 2008); Bazán and Miguel (2009); Bazán et al. (2010); Miguel (2011a,b); Gil-Albarova et al. (2013); Miguel (2013b,a, 2014)

Contents

List of Figures

Abbreviations

ICT	**I**nformation and **C**ommunication **T**echnologies
IS	**I**nformation **S**ecurity
CSCL	**C**omputer **S**upported **C**ollaborative **L**earning
LMS	**L**earning **M**anagement **S**ystems
SCLMS	**S**ecure **C**ollaborative **L**earning **M**anagement **S**ystems
PKI	**P**ublic **K**ey **I**nfrastructure
UOC	**O**pen **U**niversity of **C**atalonia
MOOC	**M**assive **O**nline **O**pen **C**ourse
NN	**N**eural **N**etwork

Chapter 1

Introduction

In this chapter we introduce and motivate the need for the research conducted in this thesis as well as discuss on the relevance of the resulting thesis' contributions to security in CSCL. To this end, we first introduce and justify. To this end, in Section 1.1 we introduce the thesis' object of research. In Section 1.2, the antecedents and current state of the research scope under study are presented. Section 1.3 presents the main challenges of this thesis' work in terms of research objectives and research questions. In order to determine the most adequate methodological approach to achieve this research objectives. Section 1.4 discusses on the most relevant aspects regarding research methodologies that conducted this thesis. Finally, the structure of the rest of this thesis report is presented in Section 1.5.

1.1 Statement of the Problem

Information and Communication Technologies (ICT) have been widely adopted and exploited in most of educational institutions in order to support e-Learning through different learning methodologies, ICT solutions and design paradigms. Over the past decade, Computer-Supported Collaborative Learning (CSCL) has become one of the most influencing learning paradigms devoted to improve teaching and learning with the help of modern information and communication technology (Koschmann, 1996). In order to support CSCL implementation, many LMS have appeared in the marketplace and the e-Learning stakeholders (i.e. e-Learning designers and managers, tutors and students)

are increasingly demanding new requirements. Among these requirements, Information Security (IS) is a significant factor involved in CSCL processes deployed in LMSs, which determines the accurate development of CSCL activities. However, according to Weippl (2006); Eibl (2010) CSCL services are usually designed and implemented without much consideration of security aspects.

The lack of security in e-Learning is also supported by practical and real attacks in ICT. As a matter of fact. Recent attack reports (CSO Magazine et al., 2011; Trustwave, 2014) have demonstrated a significant amount of real-life security attacks experimented by organizations and educational institutions. The CyberSecurity Watch Survey is a co-operative survey research conducted by reputed companies and educational institutions (CSO Magazine et al., 2011). This report reveals that security attacks are a reality for most organizations: 81% of respondents' organizations experienced a security event (i.e. an adverse event that threatens some security aspect).

Since LMS are software packages, which integrate tools that support CSCL activities, technological vulnerabilities have to be considered, recent security reports (CSO Magazine et al., 2011; Trustwave, 2014) have shown how web application servers and database management systems, which usually support LMS infrastructure, are deployed with security flaws. Dealing with more technological details related to LMSs, the Trustwave Global Security Report shows how web application servers and database management systems are deployed with security vulnerabilities (Trustwave, 2014). Moreover, potential LMS attacks can be studied by analysing their specific security vulnerabilities, for instance, in Moodle Security Announcements Moodle (2012), 49 serious vulnerabilities were reported in 2013.

Regarding security in educational institutions in the scope of the Spanish universities, the RedIRIS Computer Emergency Response Team is aimed to the early detection of security incidents affecting their affiliated institutions. As stated in Equipo de Serguridad de RedIRIS (2013), the total amount of incidents received was 10,028, and this value represents an increase of 74.15% compared to the previous year. In the same context, in Píriz et al. (2013), the authors stated that only 17% of the Spanish universities have launched the application of the Spanish National Security Framework and only 18% of students use digital certificates. Although it might seem that these plans and

initiatives are related to security in e-Learning, they are actually focused on secure e-Administration and management. In contrast, e-Learning security, which can determine these management processes, is not usually considered. For instance, a student who is able to obtain a course certificate following advanced security techniques, such as digital signature, the same security level is not required when the student is performing on-line assessment activities.

One of the key strategies in information security is that security drawbacks cannot be solved with technology solutions alone (Dark, 2011). To date, even most advances security technological solutions, such as Public Key Infrastructures (PKI) have drawbacks that impede the development of complete and overall technological security frameworks. Even most advanced PKI solutions have vulnerabilities that impede the development of a highly secure framework. For this reason, this proposal suggests to research into enhancing technological security models with functional approaches.

Among functional approaches, trustworthiness analysis, modelling, assessment and prediction methods are suitable in the context of CSCL. Trustworthiness can be considered as a suitable functional factor in CSCL because most of trustworthiness models are based on peer-to-peer interactions (Marsh, 1994; Bernthal, 1997) and CSCL is closely related to students' interactions. Although some trustworthiness methods have been proposed and investigated, these approaches have been little investigated in CSCL with the aim to enhance security properties. Therefore, this thesis proposes to conduct research on security in CSCL by enhancing technological security solutions with trustworthiness, through experimenting methods, techniques and trustworthiness models, eventually arranged in a trustworthiness methodology approach for collaborative e-Learning.

Further security applications based on trustworthiness, additional CSCL enhancements related to pedagogical factors can be considered. According to Hussain et al. (2009) the existence of trust reduces the perception of risk, which in turn improves the behaviour in the interaction and willingness to engage in the interaction. In the context of CSCL, interactions between students are one of the most relevant factors in learning performance. Therefore, trustworthiness is directly related to CSCL and can enhance the performance of collaborative learning activities. In contrast, information security can encourage and endorse trustworthiness, but IS does not directly endow learning enhancement. Another significant difference between information security and trustworthiness, with respect to

CSCL, is the dynamic nature of trustworthiness (Carullo et al., 2013). Students' behaviour is dynamic and it evolves during the CSCL process. Whilst information security is static regarding students behaviour, trustworthiness also evolves and its assessment can be adapted to students' and group' behaviour changes.

A CSCL activity is a general concept that can involve very different cases, actors, processes, requirements and learning objectives in the complex context of e-Learning (Dillenbourg, 1999). To alleviate this complexity we limit our application scope in specific CSCL activities, namely, on-line collaborative assessment (collaborative e-assessment). General e-assessment processes offer enormous opportunities to enhance student's learning experience, such as delivering on-demand tests, providing electronic marking and immediate feedback on tests (Apampa, 2010). In higher education, e-assessment is typically employed to deliver formative tests to the students. An e-assessment is an e-exam with most common characteristics of virtual exams, which are reported on unethical conduct occurring during e-learning exam taking (Levy and Ramim, 2006). In this thesis, we endowed collaborative e-assessment activities with trustworthiness assessment and prediction to enhance user security requirements.

The topics discussed so far are addressed to improve CSCL activities security with trustworthiness models. In addition to the considerations related to security, CSCL and trustworthiness, we actually need to incorporate analysis and visualization peer-to-peer systems into the security model with the aim of presenting and informing tutors regarding the results of peer-to-peer's trustworthiness.

To sum up, the target of this research is an e-Learning system formed by collaborative activities developed in a LMS. The system has to provide security support to carry out these activities and to collect trustworthiness data generated by learning and collaboration processes. Both technological frameworks and on-line collaborative learning are in line with the e-Learning strategies developed in many educational institutions. In particular, our real e-Learning context of the UOC develops full on-line education based on collaborative learning activities. Following this institutional view, information security becomes an essential issue to be considered in order for distance universities to develop secure on-line assessment processes and activities, which can conduct to grades, certificates and many types of evaluation models. The research conducted in this thesis goes to this direction and provides solid answers to the formulated research questions.

1.2 State of the Art

The antecedents and current state of the research scope under study are classified into the following topics: CSCL and e-Assessment, LMS, Information Security, Security Dimensions, Trustworthiness, Prediction, and peer-to-peer visualization. Main works in the literature on these topics are presented and analysed in this section highlighting the relations between the topics presented.

1.2.1 On-line Collaborative Learning and e-Assessment

Mature research has shown that CSCL is one of the most influencing e-Learning paradigms devoted to improve teaching and learning (Koschmann, 1996; Dillenbourg, 1999). CSCL refers to instructional methods where students are encouraged to work together on learning activities. As an example, project-based collaborative learning proves to be a very successful method to that end (Dillenbourg, 1999). Therefore, CSCL applications aim to create virtual collaborative learning environments where e-Learning students cooperate with each other in order to accomplish a common learning goal. CSCL provides support to three essential aspects, namely, coordination, collaboration and communication (Caballé, 2008).

The representation, analysis and modelling of group activity interactions are relevant processes in CSCL design, supporting evaluation and e-assessment activities in on-line collaborative learning environments (Koschmann, 1996). Collaborative learning assessment requires a broad perspective about learning and the involved processes (Erwin, 1992). Assessment processes have a significant effect on collaborative learning because they engage learners through accountability and constructive feedback (Daradoumis et al., 2006). Assessment is even more important in on-line collaborative learning environments because of the lack of real interactivity and of the feeling of social isolation of the learners (Dillenbourg, 2003). However, in order to design a coherent and efficient assessment system for collaborative learning it is necessary to design an enriched learning experience that predisposes the feedback and awareness in the group. A complex set of simultaneously applied assessment approaches, each reinforcing and/or complementing the other is the main tool to enhance collaborative learning interaction amongst group members (Strijbos et al., 2006; Daradoumis et al., 2006).

Among CSCL activities, this thesis is focused on e-assessment collaborative peer-to-peer evaluation processes and on-line collaborative activities which form e-assessment evaluation components. In Levy and Ramim (2006) the authors discussed on how unethical conduct during e-learning exam taking may occur by circumventing agreed rules, or gaining unfair advantages. These cases are closely related to e-assessment regarding anomalous evaluation processes.

1.2.2 Learning Management Systems

CSCL processes are supported in integrated LMS, according to Caballé and Feldman (2008), LMS is a broad term used for a wide range of systems that organize and provide access to on-line learning services for students, teachers, and administrators. These services usually include access control, provision of learning content, communication tools, and organizations of user groups. In other words, LMS are software components to enable the management of educational content and also integrate tools that support most of CSCL needs, such as discussion forums, chat, collaborative workspaces and repositories, and so on. Over the last years, a great amount of full-featured Web-based LMS have appeared in the marketplace (Besimi et al., 2009; Von Solms, 2010), offering designers and instructors generic, powerful user-friendly layouts for the easy and rapid creation and organization of courses and activities, which can then be customized to the tutor's needs, learners' profile and specific pedagogical goals. Since LMS are software components, information security is a relevant factor that has to be considered.

1.2.3 Information Security

Information Security (IS) in ICT can be defined as a combination of properties, which are provided by security services (Harris, 2002; Parker, 2002). The first security properties approach is the classic CIA triad that defines the three main targets of information security services: confidentiality, integrity and availability (Harris, 2002). In addition, in Parker (2002) an extension to this model is proposed including additional elements, namely, possession or access control, authenticity, and utility. However, other authors (Cheswick et al., 2003) explain, considering technological factors, that due to all general software has bugs, security software also has bugs. Finally, even though security properties defined in Parker (2002) could be taken as a first reference, since the model

proposed is not completely reliable, it is necessary to offer additional facilities, such as audit service and failure control in order to reduce the effects and negative consequences of security vulnerabilities. Hence absolute security does not exist.

Nowadays, ICT solutions based on Public Key Infrastructure (PKI) models, are available to offer technological implementations of services which ensure the security issues that have been described and required in LMSs (Raina, 2003). PKI, simply defined, is an infrastructure that allows for the creation of a trusted method for providing privacy, authentication, integrity, and non-repudiation in communications between two parties. Since 1999, PKI related standards and specifications are available, namely, the Internet X.509 PKI (PKIX) defined in IETF (2011) was developed with the aim of building Internet standards to support a pervasive security infrastructure whose services are implemented and delivered using PKI techniques.

Finally, holistic approaches of IS need to be considered (Mwakalinga et al., 2009) involving different areas, such as legal aspects and privacy legislation, secure software development, networking and secure protocols, information security management systems, standards and methodologies, certification organizations, and security testing methods and tools. These approaches have introduced more complex security properties, such as authorship or non-repudiation.

1.2.4 Security Dimensions

In order to address further technological security approaches, some authors (Schneier, 2003; Dark, 2011) have considered IS as a research topic beyond ICT. In Schneier (2003) the author stated that security is both a feeling and a reality. On one hand, reality of security is mathematical based on the probability of different risks and the effectiveness of different countermeasures. On the other hand, security is also a feeling, based on psychological reactions to both risks and countermeasures.

Moreover, absolute security does not exist and any gain in security always involves trade-offs between risk, losses, and gains (Cheswick et al., 2003). Even as it is concluded in West (2008), all security is a trade-off. This approach is very relevant in the context of this research because it is based on a hybrid security system in which technological overall solutions have to be managed beyond ICT. Moreover, in Dark (2011) the authors

discussed that problems encountered in ensuring modern computing systems cannot be solved with technology alone. Instead, IS design requires an informed, multidisciplinary approach.

Therefore, the problem of IS in CSCL is tackled with a functional approach which combine ICT security solutions with functional models, namely, trustworthiness methods and techniques in CSCL.

1.2.5 Trustworthiness

Most of trustworthiness models in the literature are related to business processes, network services and recommendation systems (Hussain et al., 2007). However, the key concept of these works is interaction between agents, that is, the same topic studied in CSCL, where agents are students and the students' interactions and trustworthiness among them are considered.

According to Gambetta (1988) there is a degree of convergence on the definition of trustworthiness, which can be defined as follows: trustworthiness is a particular level of the subjective probability with which an agent A assesses another agent AA (or group of agents). This assessment requires that the agent AA will perform a particular action, before the agent A can monitor such action.

Regarding trustworthiness and e-Learning, according to Liu and Wu (2010), a trustworthy e-Learning system is a learning system, which contains reliable serving peers and useful learning resources. From these definitions, it can be claimed that trustworthiness is closely related to both students' interactions and students' actions in CSCL. Moreover, it can be considered that trustworthiness models are focused on two different dimensions, that is, trustworthiness assessment and prediction. To establish the difference between trustworthiness assessment and prediction, in Raza et al. (2012) it is stated that trust prediction, unlike trust assessment, deals with uncertainty as it aims to determine the trust value over a period in the future.

1.2.5.1 Trustworthiness Assessment

Trustworthiness assessment is a foremost step in trustworthiness prediction. Hence, we need to review how trustworthiness can be assessed and which are the factors involved in its quantitative study. In Dai et al. (2008) a data provenance trustworthiness model is proposed This model takes into account factors that may affect trustworthiness assessment. Based on these factors, the model assigns trustworthiness scores to both data and data providers (i.e the two agents involved in the data provenance trustworthiness assessment model).

Moreover, we have to consider factors that may affect trustworthiness assessment when students are developing CSCL activities. In this sense, in Bernthal (1997) the author design a survey to explore interpersonal trust in work groups, identifying trust-building and trust-reducing behaviours ranked in order of importance. These behaviours can be used as trustworthiness factors, which can assess trustworthiness in CSCL activities.

Although trustworthiness levels can be represented as a combination of trustworthiness factors, in order to build these levels we also have to consider trustworthiness rules and characteristics. According to Liu and Wu (2010) there are different aspects of considering on trustworthiness, different expressions and classifications of trustworthiness characteristics. In essence, we can summarize these aspects defining the following rules: (i) Asymmetry, A trust B is not equal to B trust A; (ii) Time factor, trustworthiness is dynamic and may evolve over the time; (iii) Limited transitivity, if A trusts C who trusts B then A will also trust B, but with the transition goes on, trust will not absolutely reliable; (iv) Context sensitive, when context changes, trust relationship might change too.

1.2.5.2 Predicting Trustworthiness

Trustworthiness predictions models, to the best of our knowledge, have been little investigated in the context of e-assessment, even in a general prediction scope. The existing literature suggests that the term trust prediction is used synonymously and interchangeably with the trustworthiness assessment process (Raza et al., 2012).

Several studies investigating trustworthiness prediction were carried out with neural networks (Raza et al., 2012; Zhai and Zhang, 2010; Song et al., 2004). In Raza et al.

(2012), the authors propose the use of neural networks to predict the trust values for any given entities. The neural networks are considered one of the most reliable methods for predicting values (Raza et al., 2012).

In Song et al. (2004); Zhai and Zhang (2010), the authors stated that trustworthiness prediction with the method of neural network is feasible. The experiments presented in Zhai and Zhang (2010) confirm that the methods with neural networks are effective to predict trustworthiness. The work presented in Song et al. (2004) proposes a novel application of neural network in evaluating multiple recommendations of various trust standards. These cases are closely related to e-assessment regarding anomalous assessment processes as well as integrity and identity security properties.

Although we tackle the problem of predicting trustworthiness with neural network approaches, there exists other trustworthiness models without neural networks methods (Flanagin and Metzger, 2013; Liu and Datta, 2011), such as similarity approaches.

1.2.6 Analysis and Visualization of Peer-to-peer Models

In recent years, there has been an increasing amount of literature on complex networks. In Boccaletti et al. (2006) the authors review the major concepts and results recently achieved in the study of the structure and dynamics of complex networks, and summarize the relevant applications of these ideas in many different disciplines. On the one hand, scientists have to cope with structural issues, revealing the principles that are at the basis of real networks. On the other hand, many relevant questions arise when studying complex networks' dynamics, such as learning how the nodes interact through a complex topology can behave collectively (Boccaletti et al., 2006). In our context, we generate a network structure by designing peer-to-peer e-assessment components and the behaviour of the students is analysed in terms of trustworthiness for security in CSCL purposes.

Regarding social networks visualization and network graphs, there exists several software systems that cope with complex analysis requirements in social networks. According to Ackland et al. (2011) there exist many network analysis and visualization software tools, which are available to collect, analyse, visualize, and generate insights from the collections of connections formed from billions of messages, links, posts, edits, uploaded photos and videos, reviews, and recommendations. Among them, NodeXL is an open

source software tool, especially designed to facilitate learning the concepts and methods of social network analysis with visualization as a key component (Smith et al., 2009).

According to Kudelka et al. (2010), the analysis of social networks is especially concentrated on uncovering hidden relations and properties of network members. As stated by the authors in Kudelka et al. (2010), a visualization of social network is a very important part of the whole systems network architecture. The visualization tool can quickly provide relevant insight into the network structure, its vertices and their properties. In the context of e-assessment, we can apply network analysis and visualization to assessment goals, such as anomalous students behaviour. In Kudelka et al. (2010); Horak et al. (2011), the authors propose an on-line analysis tool called Forcoa.NET, which is focused on the analysis and visualization of the co-authorship relationship based on the intensity and topic of joint publications.

1.3 Objectives

The main challenge of this thesis work is to build an innovative trustworthiness methodological approach to enhance information security in collaborative e-Learning. To this end, we defined the following objectives:

O1 Define a security CSCL model.

To define a security model based on IS properties and trustworthiness intended to analyse security in CSCL activities by considering specific security requirements.

O2 Trustworthiness methodology.

To build a comprehensive trustworthiness methodology offering a guideline for the design and management of CSCL activities based on trustworthiness. This objective is composed by two sub-objectives:

O2.1 To build e-assessment peer-to-peer activities based on the trustworthiness methodology proposed.

O2.2 To propose peer-to-peer visualizations methods to manage security e-Learning events.

O3 Trustworthiness assessment and prediction.

To provide trustworthiness-based decision information in order to discover anomalous student's behaviour that compromises the security through trustworthiness assessment and prediction.

O4 Design secure CSCL activities and e-assessment.

From the security model defined in O1, the methodology built in O2 and considering trustworthiness decision information O3, to design secure CSCL activities based on trustworthiness approaches in terms of LMS components.

O5 Experimentation and validation.

To develop experimental pilots with the LMS components O4, trustworthiness assessment and prediction O3 and the trustworthiness methodology O2, with the aim to validate the enhancement of IS in CSCL activities.

Further theoretical and design objectives, experimental activities (O5) were conducted in real on-line courses. In the real learning context of the UOC, the collaborative learning processes are an essential part of its pedagogical model. Since this paradigm is massively applied to support UOC courses, security requirements can be widely analysed in this scenario. Therefore, the theoretical findings are to be tested, evaluated and verified by experimental pilots incorporated as a part of several real courses of this university.

The above objectives have motivated research on security in on-line collaborative learning by enhancing technological security solutions based on trustworthiness. The ultimate aim is to build a solid trustworthiness methodology approach for CSCL devoted to offer secure collaborative e-assessment for students, tutors and managers.

1.4 Research Methodology

In order to determine the most adequate methodological approach to achieve this research objectives, this thesis followed the research methodology: Design Science Research Process (DSRP) for Information Systems (IS) (Peffers et al., 2006; Hevner et al., 2004; Akker, 1999). Based on DSRP-IS, the following main research phases were considered:

1. Problem identification and motivation.

   Summarized in Section 1.1.

2. Objectives of a solution.

   Presented in Section 1.3.

3. Design and development.

   Presented in the rest of this section.

4. Demonstration.

   The demonstration Phase is presented in Chapter 3.

5. Evaluation.

   The evaluation Phase is presented in Chapter 3.

6. Communication.

   The thesis' contributions have been published in the conference papers and journals presented in this report (see Chapter 2 and List of Research Contributions).

Although we followed these research phases, the specific characteristics of the thesis research required to adapt some features. In particular, the Phase 3 (i.e. Design and development) was adapted to our specific methodological requirements for the design of trustworthiness CSCL components:

1. Building Trustworthiness Components integrated into the design of secure collaborative learning activities. This phase was divided into the following analysis and design tasks:

   (a) To determine security properties model.

   (b) To analyse and model students' interaction and trustworthiness.

   (c) To model security properties and students' interaction in CSCL activities.

   (d) To endow the collaborative activity with security and trustworthiness.

   (e) To design collaborative learning components.

   (f) To build students' trustworthiness profiles.

   (g) To define research instruments for data trustworthiness collection.

2. Trustworthiness analysis and data processing based on trustworthiness modelling:

   (a) To define trustworthiness modelling concepts, techniques and measures.

(b) To build normalization methods.

(c) To propose parallel processing techniques to speed and scale up the structuring and processing of basic data.

3. Trustworthiness assessment and prediction to detect anomalous students' behaviour and refine the design process:

(a) Analysis of the time factor in trustworthiness.

(b) To evaluate methods intended to predict and assess trustworthiness.

(c) Validation process to filter anomalous cases, to compare results that represent the same information from different sources, and to verify results.

(d) Management of trustworthiness decision information, security events, and anomalous students' behaviour.

(e) To offer peer-to-peer analysis and visualization tools to support the detection process of anomalous cases.

Although these phases were developed sequentially between each phase, the overall process formed by these three phases, was considered as a design cycle. Each cycle allowed enhancing the collaborative learning activities from the results (i.e. trustworthiness decision information) retrieved from the previous cycle.

Finally, the methodology for validating and evaluating each experimental pilot follows the APA guidelines for empirical studies American Psychological Association (2010).

1.5 Structure of the Thesis

The rest of this report is structured as follows. In Chapter 2 we present a summary of the three main contributions of this thesis work. The main conclusions and research results obtained as well as highlights future directions of research are presented in Chapter 3. Finally, Appendix A includes the program committees where the candidate has participated or will participate in the near future.

Chapter 2

Contributions of The Thesis

A complete copy of the three main contributions of this thesis are included in this chapter. The candidate is the first author of all the main publications of this thesis, both these main contributions and the conference papers presented in this thesis report (see List of Research Contributions).

The candidate's participation was also essential in the development of technological components described in the publications, as well as all the design, pilots and prototypes developed for the real on-line courses.

The summary, key conclusions and main research goals of these contributions are presented in Chapter 3.

2.1 Security in Online Web Learning Assessment. Providing an Effective Trustworthiness Approach to Support e-Learning Teams

Miguel, J., Caballé, S., Xhafa, F., and Prieto, J. (2015a). Security in online web learning assessment. providing an effective trustworthiness approach to support e-learning teams. *World Wide Web Journal (WWWJ)*. Springer. doi:10.1007/s11280-014-0320-2. IF: 1.623, Q1: 20/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE)

World Wide Web
DOI 10.1007/s11280-014-0320-2

Security in online web learning assessment

Providing an effective trustworthiness approach to support e-learning teams

Jorge Miguel · Santi Caballé · Fatos Xhafa · Josep Prieto

Received: 1 September 2014 / Revised: 30 November 2014 / Accepted: 22 December 2014

Abstract This paper proposes a trustworthiness model for the design of secure learning assessment in on-line web collaborative learning groups. Although computer supported collaborative learning has been widely adopted in many educational institutions over the last decade, there exist still drawbacks which limit their potential in collaborative learning activities. Among these limitations, we investigate information security requirements in on-line assessment, (e-assessment), which can be developed in collaborative learning contexts. Despite information security enhancements have been developed in recent years, to the best of our knowledge, integrated and holistic security models have not been completely carried out yet. Even when security advanced methodologies and technologies are deployed in learning management systems, too many types of vulnerabilities still remain opened and unsolved. Therefore, new models such as trustworthiness approaches can overcome these lacks and support e-assessment requirements for e-Learning. To this end, a holistic security model is designed, implemented and evaluated in a real context of e-Learning. Implications of this study are remarked for secure assessment in on-line collaborative learning through effective trustworthiness approaches.

Keywords Trustworthiness · E-Assessment · Information security · Collaborative learning

J. Miguel (✉) · S. Caballé · J. Prieto
Department of Computer Science, Multimedia, and Telecommunication,
Open University of Catalonia, Barcelona, Spain
e-mail: jmmoneo@uoc.edu

S. Caballé
e-mail: scaballe@uoc.edu

J. Prieto
e-mail: jprieto@uoc.edu

F. Xhafa
Department of Languages and Informatic Systems, Technical University of Catalonia,
Barcelona, Spain
e-mail: fatos@lsi.upc.edu

Published online: 25 January 2015

1 Introduction

Computer-Supported Collaborative Learning (CSCL) has been widely adopted in many educational institutions over the last decade. Among these institutions, the Open University of Catalonia[1] (UOC) develops on-line education based on continuous evaluation and collaborative activities.

Although on-line assessments (e-assessments) in both continuous evaluation and collaborative learning have been widely adopted in many educational institutions over the last years, there exist still drawbacks which limit their potential. Among these limitations, we investigate information security requirements in assessments which may be developed in on-line collaborative learning contexts.

Despite information security technological enhances have also been developed in recent years, to the best of our knowledge, integrated and holistic security models have not been completely carried out yet. Even when security advanced methodologies and technologies are deployed in Learning Management Systems (LMS), too many lacks still remain opened and unsolved. Therefore, as new models are needed, in this paper we propose a trustworthiness approach based on hybrid evaluation which can complete these lacks and support e-assessments requirements.

The paper is organized as follows. Section 2 shows the background about security in e-Learning as well as our research already done with respect to trustworthiness and security in e-assessment. Section 3 reviews the main factors, classification and security issues involved in security in e-assessments and we discussed that security improvements in e-assessments cannot be reached with technology alone; to fill this drawback, in Section 4, we extend our security model with the study of the trustworthiness dimension. Once studied trustworthiness factors and rules and presented our previous work, in Section 5 we describe a model based on trustworthiness applied to e-assessments. In Section 6, we conduct our research to peer-to-peer e-assessment developed in a real on-line course and by developing a statistical and evaluation analysis for the course collected data. Finally, Section 7 concludes the paper highlighting the main ideas discussed and outlining ongoing and future work.

2 Security in e-learning background

Since 1998, information security in e-Learning has been considered as an important factor in e-Learning design. Early research works about these topics [7] are focused on confidentiality and these privacy approaches can be found in [13]. Despite the relevance of privacy requirements in secure e-Learning, information security does not serve for privacy services only. Indeed, in many works [6, 23], security in e-Learning has been treated following more complex analysis and design models.

In [23] the author argues that security is an important issue in the context of education. Security is mainly an organizational and management issue and improving security is an ongoing process in e-Learning. This proposal is the first approach in which information security is applied to LMS as a general key in e-Learning design and management.

[1]The Open University of Catalonia is located in Barcelona, Spain. The UOC offers distance education through the Internet since 1994. Currently, about 60,000 students and 3,700 lecturers are involved in over 8,300 on-line classrooms from about 100 graduate, post-graduate and doctorate programs in a wide range of academic disciplines. The UOC is found at http://www.uoc.edu

Furthermore, in [6] it is presented how security in e-Learning can be analyzed from a different point of view, that is, instead of designing security, the author investigates threats for e-Learning and then, several recommendations are introduced and discussed in order to avoid detected threats. On the other hand, more specific security issues in secure e-Learning have been investigated (e.g. virtual assignments and exams, security monitoring, authentication and authorization services). These works have been summarized in [10–13].

So far we have discussed on the security design in e-Learning from a theoretical point of view. However, some authors argue we actually need to understand attacks in order to discover those relevant security design factors and figure out how security services must be designed [5]. Researchers have already conducted many efforts proposing taxonomies of security attacks. In [24], through analyzing existing research in attack classification, a new attack taxonomy is constructed by classifying attacks into dimensions. This paper also offers a complete a complete and useful study examining existing proposals. Nevertheless, since attacks taxonomies might be applied to cover each kind of attack, which might occur in LMS, they are not closely related to security design in e-Learning. In order to fill this gap, in [13], we have proposed an alternative approach which associate attacks to security design factors.

We now extend the background about security in e-Learning by analyzing real-life security attacks and vulnerabilities, which could allow attackers to violate the security in a real context. In this sense, several reports are found, which justify the relevance of security attacks during the last two years. In particular, the study presented in [2] uncovered that security attacks are a reality for most organizations: 81 % of respondents' organizations experienced a security event (i.e. an adverse event that threatens some aspect of security). Finally, we can consider specific LMS real software vulnerabilities. Moodle is an Open Source LMS which is massively deployed in many schools and universities. In Moodle Security Announcements [2], 40 serious vulnerabilities have been reported in 2013.

In previous research [10–13] we have argued that general security approaches do not provide the necessary security services to guarantee that all supported learning processes are developed in a reliable way. The rest of this section presents our work already done and our research results obtained at the time of this writing regarding analysis and security design in CSCL, trustworthiness and e-assessment and a trustworthiness methodology proposal.

We have investigated how to enhance CSCL security in terms of security analysis and design. To this end, we have analysed security properties models, how to model students' interaction and trustworthiness, and how security properties and students' interaction are involved in CSCL activities. These goals and research results are summarized in the following list:

- Security requirements in CSCL. In [11] it is argued that current e-Learning systems supporting on-line collaborative learning do not sufficiently meet essential security requirements and this limitation can have a strong influence in the collaborative learning processes.
- Design of secure CSCL systems. In [10] the problems caused in collaborative learning processes by the lack of security are discussed and the main guidelines for the design of secure CSCL systems are proposed to guide developers to incorporate security as an essential requirement into the collaborative learning process.

[2] https://moodle.org/mod/forum/view.php?f=996

- Security requirements in mobile learning. In [12] it is presented an overview of secure LMSs, inspecting which are the most relevant factors to consider, and connecting this approach to specific aspects for mobile collaborative learning. Then, real-life experience in security attacks in mobile learning are reported showing a practical perspective of the learning management system vulnerabilities. From this experience and considerations, the main guidelines for the design of security solutions applied to improve mobile collaborative learning are proposed.
- Security requirements in MOOCs. In [13] it is investigated the lack of provision of IS to MOOC, with regards to anomalous user authentication, which cannot verify the actual students identity to meet grading requirements as well as satisfy accrediting institutions. In order to overcome this issue, it is proposed a global user authentication model called MOOC-SIA.

Once security and CSCL issues have been analysed, we have focused our research work on trustworthiness analysis and data processing based on trustworthiness modelling in order to define trustworthiness modelling concepts (i.e. techniques and measures). The aim is to build normalization methods and propose parallel processing techniques to speed and scale up the structuring and processing of basic data. These objectives are related to the design of secure learning objects, trustworthiness assessment and prediction, and the development of pilots for validation processes. This work has produced the following research results:

- Trustworthiness model. In [15] a trustworthiness model for the design of secure learning assessment in on-line collaborative learning groups is proposed. To this end, a trustworthiness model is designed in order to conduct the guidelines of a holistic security model for on-line collaborative learning through effective trustworthiness approaches.
- Parallel processing approach. In [14] it is proposed a trustworthiness-based approach for the design of secure learning activities in on-line learning groups. The guidelines of a holistic security model in on-line collaborative learning through an effective trustworthiness approach are presented. As the main contribution of this paper, a parallel processing approach, which can considerably decrease the time of data processing, is proposed thus allowing for building relevant trustworthiness models to support learning activities even in real-time.
- Trustworthiness normalization methods. In [19] an approach to enhance information security in on-line assessment based on a normalized trustworthiness model is presented. In this paper, it is justified why trustworthiness normalization is needed and a normalized trustworthiness model is proposed by reviewing existing normalization procedures for trustworthy values applied to e-assessments. Eventually, the potential of the normalized trustworthiness model is evaluated in a real CSCL course.
- Trustworthiness prediction. In [18] previous trustworthiness models are endowed with prediction features by composing trustworthiness modelling and assessment, normalization methods, history sequences, and neural network-based approaches. In order to validate our approach, a peer-to-peer e-assessment model is presented and carried out in a real on-line course.

The next phase of our research on security in e-Learning based on trustworthiness has been focused on building a trustworthiness methodology offering a guideline for the design and management of secure CSCL activities based on trustworthiness assessment and prediction to detect security events and evidences. In [17] the need of trustworthiness models as a functional requirement devoted to improve information security is justified. A methodological approach to modelling trustworthiness in on-line collaborative learning were proposed.

This proposal aims at building a theoretical approach to provide e-Learning designers and managers with guidelines for incorporating security into on-line collaborative activities through trustworthiness assessment and prediction.

Finally, we have endowed our trustworthiness approaches with the concept of students' profile and collective intelligence features. In [16] we have discovered how security can be enhanced with trustworthiness in an on-line collaborative learning scenario through the study of the collective intelligence processes that occur on on-line assessment activities. To this end, a peer-to-peer public students profile model, based on trustworthiness is proposed, and the main collective intelligence processes involved in the collaborative on-line assessments activities were presented.

To sum up, the present paper contribute to existing security solutions models by providing an innovative approach for modelling trustworthiness in a real context of secure learning assessment in on-line collaborative learning groups. The study shows the need to combine technological security solutions and functional trustworthiness measures.

3 Secure e-assessment

In this section, we present a review of the main factors, classification and security issues involved in security in e-assessments. Firstly, security properties related to e-assessments are evaluated by examining and selecting most relevant ones. Then, an assessments classification is depicted in order to analyse how e-assessments types and factors are related to previously selected security properties and. Finally, we propose a security model which extends technological security techniques adding functional requirements to secure e-assessments.

3.1 Authenticity in e-assessments

In order to determine whether or not an e-assessment is secure, both from students' as evaluators' point of view, it can be inquired if the e-assessment satisfies the following properties:

- Availability. The e-assessment is available to be performed by the student at the scheduled time and during the time period which has been established. After the assessment task, the tutor should be able to access the results to proceed to review the task.
- Integrity. The description of the e-assessment (statement of the activity, etc.) must not be changed, destroyed, or lost in an unauthorized or accidental manner. The result delivered by the student must achieve the integrity property too.
- Identification and authentication. While performing the evaluation task, the fact that students are who they claim to be must be verifiable in a reliable way. In addition, both students' outcomes and evaluation results must actually correspond to the activity that students have performed.
- Confidentiality and access control. Students will only be able to access to e-assessments that have been specifically prepared to them and tutors will access following the established evaluation process.
- Non repudiation. The LMS must provide protection against false denial of involvement in e-assessments.

Due to the difficulty of provisioning a complete secure e-assessment including all of these properties, a first approach of secure e-assessments selects a subset of properties which

can be considered as critical in evaluation context. The selected properties are identification and integrity. Integrity must be considered both as authorship as well as data integrity. Therefore, we will be able to trust an e-assessment process when identification and integrity properties are accomplished. In the context of e-assessments, with regarding to identification, students are who they claim to be when they are performing the evaluation activities (e.g. access to the statement in a test, answering a question in an interview with the evaluator, etc.). In addition, dealing with integrity and authorship, we trust the outcomes of the evaluation process (i.e. a student submits evaluation results) when the stunted is actually the author and these elements have not been modified in an unauthorized way. It is important to note that e-assessments are developed in a LMS and, since the LMS is an information system, two different items are involved in this context: processes and contents which are related to integrity and identification. Therefore, services applied to e-assessment must be considered in both a static and a dynamic way.

3.2 Assessments calassification

The scope of our research, with regarding to assessment, is the evaluation model used in UOC courses. Evaluation models used in UOC may be classified in accordance with the following factors or dimensions: (i) type of subjects; (ii) specific evaluation model; (iii) evaluation application; (iv) agents involved in the evaluation processes. Figure 1 shows factors and evaluation types.

Firstly, we have to analyse the agents who are involved in evaluations processes. The agents selected are students, tutors and the LMS, that is, students carrying out learning activities in a LMS which are assessed by tutors. In this context, we consider two types of subjects in UOC courses, a standard subject has many students in the virtual classroom and

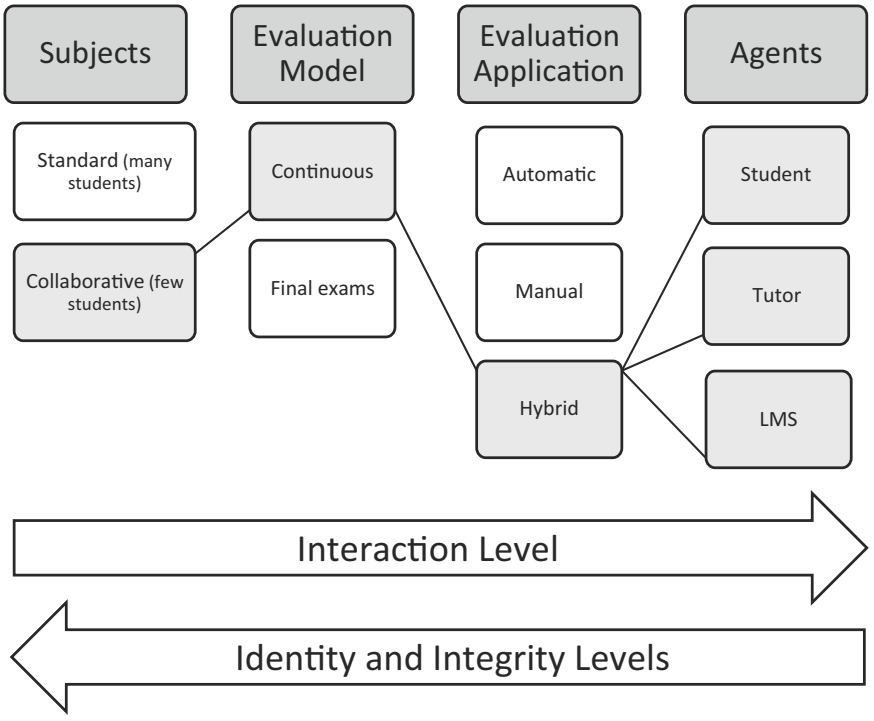

Figure 1 Evaluation types

the level of collaborative learning activities is low. On the other hand, a collaborative subject is designed following a intensive collaborative learning model which is performed by few students arranged in learning groups. Regarding these evaluation models, two different models are selected, the continuous evaluation model allows the tutors to assess the students throughout the course by evaluating each activity in the subject; in contrast, a evaluation model based on final exams focuses the evaluation processes on an assessment instrument at the end of the course.

Once the subject, evaluation and agent dimension are presented, we focus the analysis on evaluation applications. In manual evaluation methods, tutors usually participate directly and intensely in the evaluation process. This model has scalability problems but can provide better guarantees for students' identification and authorship because the degree of interaction between tutors and students is higher than in others evaluation methods. Although this statement may be true in general cases, it may not apply to all situations, that is, the interaction level does not necessarily mean that students' identification is authentic (as defined above: data integrity and authorship). On the other hand, automatic methods do not involve tutors participation (or minimal), but this model does not carry out desirable identification and integrity levels. Finally, hybrid methods are a trade-off combination which can provide a balance between the degree of interaction and security requirements. In Figure 1 it has been marked those elements which are involved in the model proposed. In the following sections, the secure e-assessment model is presented.

3.3 Technological approaches

According to [4] problems encountered in ensuring modern computing systems cannot be solved with technology alone. In order to probe this statement and to justify that it is needed to extend technological models with trustworthiness functional proposals, in this section, we are going to present a use case that illustrate how Public Key Infrastructure (PKI) does not completely guarantee security requirements. The example use case is defined as follows:

The e-assessment is an e-exam with most common characteristics of virtual exams. For further information, in [8] it is discussed how unethical conduct during e-Learning exam taking may occur and it is proposed an approach that suggests practical solutions based on technological and biometrics user authentication.

The e-exam is synchronous and students have to access the LMS to take the description of the e-exam at the same time. The exam, which presents a list of tasks to be solved by the student. The statement is the same for all students who perform the e-exam and then, each student performs her work into a digital document with her own resources. When the student's work is finished, outcomes are delivered to the LMS before the deadline required.

Once defined this use case, we can improve security requirements using PKI based solutions, in concrete terms, digital certificates to guarantee students' identification and digital signature for outcomes integrity and authorship. Therefore, the process described above is adapted to this way:

- The student accesses the LMS identified by its digital certificate. Similarly, the LMS presents its digital certificate to the student.
- Since both LMS and student have been identified in a trust process, the student receives the description of the e-exam and begins her work.
- The student checks the built-in digital signature statement in order to validate the integrity of this element.

– When the student finishes her work in the outcomes document, the student performs the operation of digital signature (into the digital document and using her digital certificate).
– Eventually, the student's signed document will be delivered in the LMS, according to the procedure defined in the first step.

At this point we can formulate the question: can we trust this model? In other words, are those processes and elements involved in the e-exam bearing integrity and identification properties? As stated at the beginning of this section, ensuring modern computing systems cannot be solved with technology alone. Therefore, we should be able to find vulnerabilities in this technological security proposal. For instance, although the identification process based on the certificate public key (even signed and issued by a certification authority) is only able to be made by the holder of the private key (the student), we do not know if this certificate is being used by the student who we expect or if the student has sent this resource to another one. Although we can add additional technological measures such as certificate storage devices, there are ways to export these keys or have remote access to manage them. Therefore, we can conclude that the student may share their resources identification and signature.

4 Trustworthiness approaches for secure e-assessment

In the previous section we discussed that security improvements in e-assessments cannot be reached with technology alone. To fill this drawback that impedes e-assessments to deploy their potential, we review in this section trustworthiness approaches to design secure e-assessment.

4.1 Trustworthiness and security related work

In [22] it is discussed that security is both a feeling and a reality. The author points out that the reality of security is mathematical based on the probability of different risks and the effectiveness of different countermeasures. In addition, security is a feeling based not on probabilities and mathematical calculations, but on our psychological reactions to both risks and countermeasures. Since this model considers two dimensions in security and being aware that absolute security does not exist (see Section 3.3) any gain in security always involves a trade-off between technological and functional approaches. This approach is very relevant in the context of hybrid evaluation systems in which technological and trustworthiness solutions can be combined. This trade-off is proposed because, as it is concluded by the author, we need both to be and to feel secure.

Our approach providing security to e-assessments extends technological solutions and combines these services with trustworthiness models. In this context, it is also important to consider additional trustworthiness related work, even when the scope of trustworthiness models is not closely related to security in e-Learning. Next, we continue our related work study taking general trustworthiness references.

4.2 Trustworthiness factors

Beyond the overview of security and trustworthiness presented, we need to review how trustworthiness can be measured and which are the factors involved in its quantitative study.

In [3] a data provenance trust model is proposed, which takes into account factors that may affect the trustworthiness. Based on these factors, the model assigns trust scores to both data and data providers.

In our context, students and students' resources (e.g. a document, a post in a forum, etc.) can be modelled following this approach. Moreover, factors that may affect trustworthiness when students are developing collaborative learning activities must be discovered. To this end in [1], the author designs a survey to explore interpersonal trust in work groups identifying trust-building behaviours ranked in order of importance. We use these behaviours as trustworthiness factors which can measure trustworthiness in those activities that students develop in collaborative activities. The factors considered to model trustworthiness when students are performing collaborative activities are summarized in Table 1.

4.3 Trustworthiness rules and characteristics

Trustworthiness levels may be represented as a combination of trustworthiness factors. Moreover, according to [9] there are different aspects of consideration of trust and different expressions and classifications of trust characteristics. In essence, we can summarize these aspects defining the following rules: (i) Asymmetry, A trust B is not equal to B trust A; (ii) Time factor, trustworthiness is dynamic and may evolve over the time; (iii) Limited transitivity, if A trusts C who trusts B then A will also trust B, but with the transition goes on, trust will not absolutely reliable; (iv) Context sensitive, when context changes, trust relationship might change too.

The model presented in this paper is designed taking into account factors and rules which have been presented in this section. Furthermore, we define two additional concepts (trustworthiness levels and indicators) which are presented in the following sections.

Table 1 Trustworthiness factors

N	Factors and Description
	Trustworthiness Building Factors (TBF)
	Student S working in the group of students GS is building trustworthiness when...
1	S communicates honestly, without distorting any information.
2	S shows confidence in GS's abilities.
3	S keeps promises and commitments.
4	S listens to and values what GS say, even though S might not agree.
5	S cooperates with GS and looks for mutual help.
	Trustworthiness Reducing Factors (TRF)
	Student S working in the group of students GS is reducing trustworthiness when...
1	S acts more concerned about own welfare than anything else.
2	S sends mixed messages so that GS never know where S stands.
3	S avoids taking responsibility.
4	S jumps to conclusions without checking the facts first.
5	S makes excuses or blames others when things do not work out.

4.4 Evidences and signs

Trustworthiness factors are defined from the perspective of students' behaviours and, on the other hand, technological solutions cannot solve security requirements alone; in consequence, it is necessary to note that all methods discussed provide security improvements but do not completely ensure e-assessments requirements. Furthermore, neither trustworthiness nor PKI models define or manage the actions to take when the security service detects either anomalous situations or violation of the properties we have defined. Firstly we must consider that according to this fact we have to distinguish between evidences and signs. Evidence is defined as information generated by the security system in a reliable way and the evidence allows us to state that a certain security property has been violated. For example, if a process of electronic signature is wrong, we can state that the signed document does not meet the integrity property and this is an irrefutable fact regarding to mathematical properties of public and private keys involved in digital signature. On the other hand, signs allow us to assign a trustworthiness level to a system action or result. These levels are based on probabilities and mathematical calculations, in other words, potential anomalous situations are associated with probabilities.

For each type of anomalous situations detected (i.e. evidences and signs) it is necessary to define different measures. Measures which can be taken are presented below:

- Active. We act directly on the e-assessments processes. For instance, if a evidence is detected, the security service will deny access to the student and the student cannot continue with the next tasks.
- Passive. Analysis and audit. Focused on analysing the information provided by the security system without acting on the e-assessment. They may generate further actions, but the process continues as planned before the fault detection.

5 A trustworthiness model

In this section, we propose a trustworthiness model for security based on the previous elements and issues. Firstly, we identify those instruments and tools which will collect trustworthiness data. Then, a statistical analysis based on a model of trustworthiness levels is presented.

5.1 Research instruments and data gathering

Four research instruments are considered to collect users' data for trustworthiness purposes and feed our model:

- Ratings. Qualifications of objects in relation to assessments, that is, objects which can be rated or qualified by students in the LMS.
- Questionnaires. Instruments which allow us to both collect trustworthiness students' information and to discover general aspects design in our model.
- Students' reports. Assessment instrument containing questions and ratings performed by the students and reviewed by the tutors.
- LMS usage indicators. To collect students' general activity in LMS (e.g. number of documents created).

All of these research instruments are quantitative and they have been designed to collect mainly trustworthiness levels and indicators as well as assessment information. In order to manage trustworthiness data, we define the concept of trustworthiness Data Source (DS) as those data generated by the research instrument that we use to define trustworthiness levels which are presented in the following section.

5.2 Modelling trustworthiness levels, indicators and rules

We introduce now the concept of trustworthiness indicator tw_i (with $i \in I$, where I is the set of trustworthiness indicators) as a measure of trustworthiness factors. Trustworthiness factors have been presented (see Section 4.2) as those behaviours that reduce or build trustworthiness in a collaborative group and they have been considered in the design of questionnaires. For instance, a trustworthiness indicator measuring the number of messages in a forum is related to the TBF-5 (the student cooperates and looks for mutual help). Therefore, an indicator tw_i is associated with one of the measures defined in each e-assessment instrument (i.e. ratings, questionnaires, reports, etc.). Moreover, we introduce the concept of trustworthiness level Ltw_i as a composition of indicators over trustworthiness rules and characteristics. For instance, we can consider two trustworthiness indicators (tw_a and tw_b). These indicators are different, the first indicator could be a rating in a forum post and the second one could be a question in a questionnaire; but they measure the same trustworthiness building factor (e.g. TBF-1: communicates honestly, described in Table 1). Finally, trustworthiness rules R, may be compared to the group, over the time or considering the context. Considering all the above, trustworthiness indicators can be represented following these expressions:

$$tw_{a_{r,s}}, a \in \{Q, RP, LGI\}, r \in R, s \in S \tag{1}$$

where Q is the set of responses in Questionnaires, RP is the analogous set in Reports, LGI is the set of LMS indicators for each student (i.e. ratings and the general students' data in the LMS). S is the set of students in the group and R is the set of rules and characteristics (e.g. time factor). These indicators are described above when presenting research instruments.

Once trustworthiness indicators have been selected, trustworthiness levels can be expressed as follows:

$$Ltw_i = \sum_{i=1}^{n} \frac{tw_i}{n}, i \in I \tag{2}$$

where I is the set of trustworthiness indicators which are combined in the trustworthiness level Ltw_i.

Trustworthiness levels Ltw_i must be normalized. To this end, we have reviewed the normalization approach defined in [21] with regarding to support those cases in which particular components need to be emphasized more than the others. Following this approach, we previously need to define the weights vectors:

$$w = (w_1, \ldots, w_i, \ldots, w_n), \sum_{i}^{n} w_i = 1 \tag{3}$$

where n is the total number of trustworthiness indicators and w_i is the weight assigned to tw_i. Then, we define trustworthiness normalized levels as:

$$Ltw_i^N = \sum_{i=1}^{n} \frac{(tw_i \cdot w_i)}{n}, i \in I \tag{4}$$

To sum up, our trustworthiness approach allows us to model students' trustworthiness as a combination of normalized indicators using research and data gathering instruments. Regarding groups, this model may also be applied in cases with only one working group; in this scenario, all students would belong to the same group.

5.3 Statistical analysis

Following the trustworthiness model presented we need to inquire whether the variables involved in the model are correlated or not. With this purpose the correlation coefficient may be useful. Some authors have proposed several methods with regarding to rates of similarity, correlation or dependence between two variables [20]. Even though the scope of [20] is focused on user-based collaborative filtering and user-to-user similarity, the models and measures of the correlations between two items applied in this context are fully applicable in our scope. More precisely, we propose Pearson correlation coefficient (represented by the letter r) as a suitable measure devoted to conduct our trustworthiness model. Pearson coefficient applied to a target trustworthiness indicator is defined bellow:

$$r_{a,b} = \frac{\sum_{i=1}^{n} \left(tw_{a,i} - \overline{tw}_a\right)\left(tw_{b,i} - \overline{tw}_b\right)}{\sqrt{\sum_{i=1}^{n} \left(tw_{a,i} - \overline{tw}_n\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(tw_{b,i} - \overline{tw}_b\right)^2}} \tag{5}$$

where tw_a is the target trustworthiness indicator, tw_b is the second trustworthiness indicator in which tw_a is compared (i.e. similarity, correlation, anomalous behaviour, etc.), $\overline{tw}_a$ and $\overline{tw}_b$ are the average of the trustworthiness indicators and n is the number of student's provided data for tw_a and tw_b indicators.

It is important to note that if both a and b are trustworthiness indicators which have several values over the time (e.g. a question which appears in each questionnaire), they must be compared at the same point of time. In other words, it is implicit that $r_{a,b}$ is actually representing r_{a_t,b_t} where a_t is the trustworthiness indicator in time t.

In addition, this test may be applied to every trustworthiness indicator taking one of them as target indicator. To this end, we define the general Pearson coefficient applied to a target trustworthiness indicator over the whole set of indicators is defined as follows:

$$r_{a,t} = \left(r_{a,1}, \ldots, r_{a,i}, \ldots, r_{a,n-1}\right), i \in I, i \neq a \tag{6}$$

where $r_{a,i}$ is the Pearson coefficient applied to a target trustworthiness indicator is defined above and I is the set of trustworthiness indicators.

Both relation and similarity are represented by $r_{a,b}$ and r_A grouping students' responses and taking the variables at the same time. We are also interested in time factor and it may be relevant the evolution of trustworthiness indicators throughout the course. To this end, we extend previous measures, adding time factor variable:

$$r_{a,t,tt} = \frac{\sum_{i=1}^{n} \left(tw_{a_{t,i}} - \overline{tw}_{a_t}\right)\left(tw_{a_{tt,i}} - \overline{tw}_{a_{tt}}\right)}{\sqrt{\sum_{i=1}^{n} \left(tw_{a_{t,i}} - \overline{tw}_{a_t}\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(tw_{a_{tt,i}} - \overline{tw}_{a_{tt}}\right)^2}} \tag{7}$$

where t is the target point in time and tt is the reference point in time (i.e. t is compared against tt), all other variables have already been defined with this case they are instanced in two moments in the course.

Similarly, we can calculate $r_{a,t,tt}$ for each tt, and then the following indicator may be used:

$$r_{a,t} = \left(r_{a,1}, \ldots, r_{a,i}, \ldots, r_{a,n-1}\right), i \in I, i \neq a \tag{8}$$

Table 2 Trustworthiness Basic Indicators

Indicator	Description	Group by	Target/Reference
$r_{(a,b)}$	Pearson coefficient applied to a target trustworthiness indicator.	Students	tw_a and tw_b
r_a	$r_{(a,b)}$ over the set of indicators	Indicators	tw_a
$r_{(a,t,tt)}$	Pearson coefficient applied to a tw indicator throughout the course from t to tt	Time	tw_a and t
$r_{(a,t)}$	$r_{(a,t,tt)}$ over the throughout the course.	Course	tw_a

Trustworthiness indicators which have already been presented in this section are summarized in Table 2.

Since hybrid methods are considered as a suitable trade-off approach for the model, we can combine these indicators with results of manual continuous evaluation results made by the tutor. For instance, a coefficient applied to target trustworthiness indicator a is compared to a manual continuous evaluation, that is:

$$r_{a,b} = cv_t \tag{9}$$

where the second indicator b is exchanged by the value in continuous evaluation. According to this indicator, we can analyse the similarity between manuals and automatics results. Furthermore, each Pearson interpretation which has been presented until now, may be applied to continuous evaluations parameters, for instance: $r_(a, t, tt)$ where $a = cv_t$.

On the other hand, as aforementioned in the case of questionnaires, some questions, which evaluate the same trustworthiness factor, are proposed in two different ways: individual and group evaluation. Hence, students are asked about some factors related to every member in her work group and then about the group in general. In this case, we can also compare these values using Pearson correlation. Finally, trustworthiness indicators may be gathered in a trustworthiness matrix with the aim of representing the whole relationship table for each indicator:

$$R_{tw} = \begin{pmatrix} 0 & r_{tw_1,tw_2} & \cdots & \cdots & r_{tw_1,tw_n} \\ 0 & 0 & r_{tw_2,tw_3} & \cdots & r_{tw_2,tw_n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & r_{tw_{n-1},tw_n} \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix} \tag{10}$$

Indicators which have been presented in this section are studied in the analysis stage of the model. Although they are proposed as suitable options, the model is refined to select those indicators oriented to perform the best similarity and correlation evaluation model. In addition, this approach is also intended to be a prediction tool, that is, similarity facts may conduct to carry out predictions about the evaluation system and its evolution.

6 Analysis of results and evaluation

As discussed in the Section 2 with respect to trustworthiness models and bearing in mind the abstract model presented in the Section 5, there exist considerable variation regarding goals, contexts, and scopes in trustworthiness approaches. In this section, we conduct our

evaluation method on peer-to-peer e-assessment developed in a real on-line course. Our peer-to-peer e-assessment model is based on a collaborative assessment component and, in this section, we also present the design and implementation of the component including research instruments and technological tools. Finally, we conclude the section with important issues concerning processing trustworthiness levels and indicators as well as statistical analysis and interpretation.

6.1 Real on-line course features

We have carried out several studies [15, 17, 18] in our real context of e-Learning of the UOC during the Spring academic term of 2014, with the aim to experiment with specific trustworthiness and security approaches devoted to evaluate the feasibility of our trustworthiness models, tools, and methodologies. In this paper, we build and deploy our comprehensive e-assessment methodology in the real on-line course presented in [15, 17, 18], whose key features can be summarized as follows:

- Students' e-assessment was based on a manual continuous e-assessment model by using several manual e-assessment instruments.
- Manual e-assessment was complemented with automatic methods, which represented up to 20 percent of the total students overall grade.
- Taking into account below features, we implemented a hybrid e-assessment method by combining manual and automatic e-assessment methods, and the model allows us to compare results in both cases.
- 59 students performed a subjective peer-to-peer e-assessment, that is, each student was able to assess the rest of class peers in terms of knowledge acquired and participation in the class assignments.
- The course followed seven stages which were taken as time references in trustworthiness analysis. These time references allow us to compare trustworthiness evolution as well as to carry out e-assessment methods.
- Each stage corresponded to a module of the course, which had a learning component (i.e. book) that the student should have studied before developing the assessment activities of the course.

From the above methodology, we have designed the peer-to-peer e-assessment component which is presented in the next section.

6.2 Continuous assessment component

As aforementioned in Section 3.1, we used a subset of security properties for e-assessment security modelling, hence integrity and identification were selected as target security properties for the continuous assessment component. Following these security properties and after the analysis of potential students' interactions in peer-to-peer assessment activities as well as the peer-to-peer assessment possibilities, the first version of the continuous assessment component was proposed in [17, 18].

The Continuous Assessment (CA) component is formed by the following three assessment activities and procedures [18]:

1. Once the student has studied a module (M), she receives an invitation to answer a set of three questions about the current module; this is the first activity of the CA named the Module Questionnaire and denoted by Q.

2. The student does not have to answer as soon as Q is sent, because the second activity of the CA is a students' forum (F) intended to create a collaborative framework devoted to enhance responses in activity Q, in other words, Q and F activities are concurrent tasks.
3. The final activity is the core of the peer-to-peer assessment and the student has to complete a survey (P) which contains the set of responses from Q. The student has to assess each classmates' responses in Q and, furthermore, the activity of each student in the forum F is assessed. The scale used to assess both forum participation and students' responses is (A, B, C+, C-, D, and N for no answer).

The formulation of the algorithm corresponding to the e-assessment process of the CA was presented in [18] (see Algorithm 1 and also [17]).

Algorithm 1 Algorithm for the e-assessment process [18]

Require: M {the list of modules} and S {the set of students in the course}

fOr m: M **do**
$Q_m \leftarrow create\_questionnaire(m)$
$send(Q_m, S)$
$F_m \leftarrow create\_forum(m)$
$F(m) \leftarrow class\_discussion(F_m, S)$
$Q(m) \leftarrow getResponses(Q_m, S)$
$P_m \leftarrow create\_p2p\_eval(Q(m), S)$
$send(P_m)$
$P(m) \leftarrow getResponses(P_m, S)$
$e\_assessment(m)[] \leftarrow results(Q, F, P, S)$
end for
return $e\_assessment(m)[]$

6.3 Research instruments and technological tools

For the purpose of the CA implementation and deployment, a questionnaire creation function has been developed (i.e. create_questionnaire). Due to the output of the first questionnaire (see variable Q(m) in the algorithm) is the input to the peer-to-peer assessment activity (i.e. variable P_m), we can automate the assessment process for each CA. These function has been implemented as a Java class named *CreateP2P*, which includes the set of attributes and methods required to automatically generate the assessment activity P_m. The automation capabilities of the process are actually focused on the set of responses and the survey P_m manual customizations such as the text or the invitation messages.

The CA uses two survey web applications. The module questionnaire (Q) is implemented in Google Forms[3] and the peer-to-peer questionnaire (P) with LimeSurvey[4]. Due to the data exchange requirements between the two survey tools, we have selected the Coma Separate Values (CSV) format as the data exchange model. For this reason and with the aim of simplifying the implementation process we have integrated in our Java components the

[3] http://www.google.com/drive/apps.html

[4] http://www.limesurvey.org

package Super CSV[5] which offers advanced CVS features dealing with reading and writing advanced operations on lists of strings.

We have selected LimeSurvey because a high configurable export and import survey functions based on standard formats are needed. After the evaluation of several survey formats, we have selected the CSV option. The function *create_p2p_eval* has been implemented by the Java class *create_p2p_csv*, which receives a CSV responses file containing the set of responses collected by Google Forms and creates a LimeSurvey CVS survey format by converting the responses in questions for the new peer-to-peer questionnaire. The hosting support for LimeSurvey framework has been provided by the RDlab[6].

Moreover, because of the peer-to-peer and dynamic features of the questionnaire P, we need to extract assessment results in primitive and normalized e-assessment data format as presented in the following section. To this end, we have developed the Java class *Results*.

Finally, dealing with processing the Pearson correlation coefficient, we have used the statistical analysis program GNU PSPP[7].

6.4 Trustworthiness data sources, levels and indicators

Before the statistical analysis phase, we define trustworthiness data sources, indicators and levels in the context of our CA. We have defined a trustworthiness data source as those data generated by the CA that we use to define trustworthiness features presented in Section 4. Each CA (i.e. one CA per module) will manage four data sources. The first is related to the students' responses count and can be denoted with the following ordered tuple:

$$DS_{Q_C} = (M, Q, S, count) \tag{11}$$

where the questionnaire data source is defined as the total number of responses ($count$) that each student in S has answered in the questionnaire Q for the module M.

The second data source also refers to the students' responses and the DS offers each specific response:

$$DS_{Q_R} = (M, Q, S, res) \tag{12}$$

where the questionnaire data source DS_{Q_R} is defined as the response res (i.e. a student answers res to a question) that each student in S has responded regarding a specific question in Q in the module M.

The third data source refers to the participation degree in a forum. These data sources can be denoted with the following ordered tuple:

$$DS_F = (M, F, S, count) \tag{13}$$

where the forum data source DS_F is defined as the total number of posts ($count$) that each student in S has sent to a forum F regarding a specific question in Q in the module M.

Finally, we introduce a score data source as follows:

$$DS_R = (M, Q, S, SS, score) \tag{14}$$

where the responses data source denotes the score that a student (in S) has assessed a student's (in SS) response of a question in Q. Hence, S is the set of students who assess and SS is the set of students who are assessed by students in S. Although S and SS may be

[5] http://supercsv.sourceforge.net/index.html

[6] http://rdlab.lsi.upc.edu

[7] http://www.gnu.org/software/pspp/

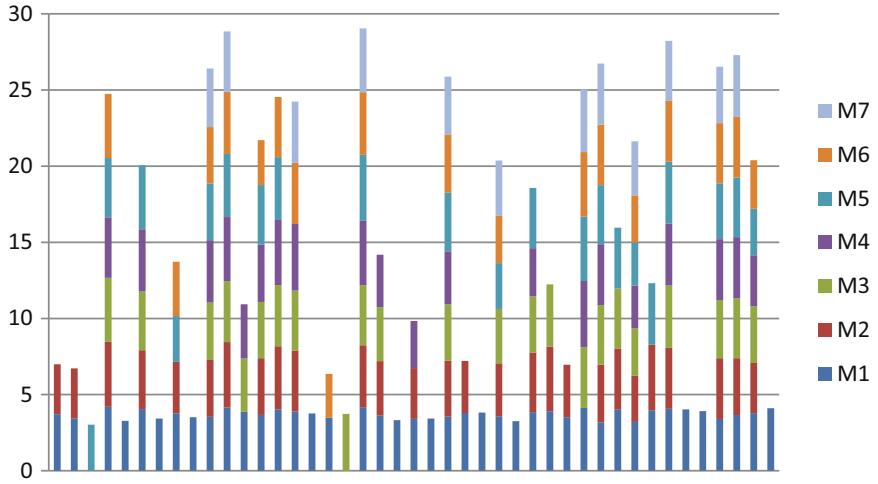

Figure 2 $L_{R,m,s}$ level for each student and module

considered as the same set of students in certain applications, they are actually considered as different sets because we permit participation in the second stage of the activity even when the student has not carried out the first one.

Tuples in DS_R are stored in a relational database table, namely MySQL[8].

Once trustworthiness data sources have been defined we define three trustworthiness levels. Following the model defined in Section 5.3, we first combine the trustworthiness indicators of each question in the module, and then the overall trustworthiness level for the student in a specific module is defined:

$$L_{R,m,s} = \sum_{i=1}^{n} \frac{(tw_i \cdot w_i)}{n}, i \in Q, w = \left(w_i = w_j\right), m \in M \tag{15}$$

where $L_{R,m,s}$ is the trustworthiness level for the student s in the module m measured by the trustworthiness indicator tw_i which considers the responses for each question in Q.

$$L_{F,m,s} = tw_{F,m}, m \in M \tag{16}$$

where $tw_{F,m}$ is the trustworthiness indicator for the responses in the collaborative forum F for the module m.

$$L_{m_i s} = \sum_{i=1}^{n} \frac{Ltw_i \cdot w_i}{n}, i \in \{L_{R,m}, L_{F,m}\}, w = \left(w_i = w_j\right), m \in M \tag{17}$$

where $L_{m_i s}$ is the overall trustworthiness level for the student s in the module m, calculated by combining the trustworthiness level for responses $L_{R,m,s}$ and the trustworthiness level for forum participation $L_{F,m,s}$.

6.5 Statistical analysis and interpretation

Here we analyse the trustworthiness levels and indicators presented in the previous section. The graph presented in Figure 2 shows the overall $L_{R,m,s}$ for each student and for each module. It is worth mentioning that students who had not participated in any CA activity

[8] http://www.mysql.com/

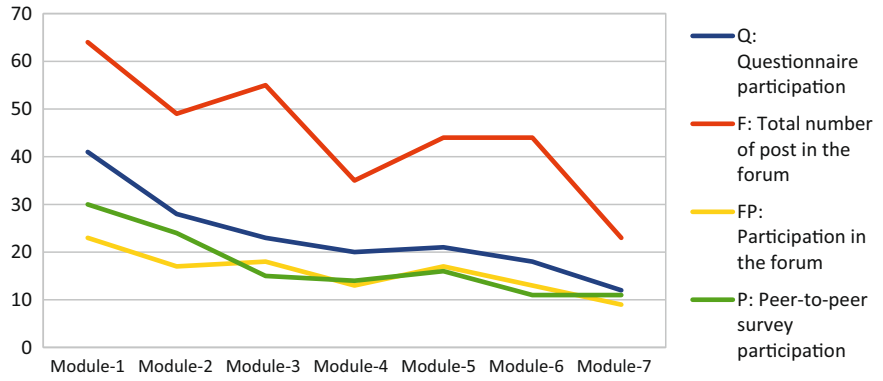

Figure 3 Students' participation evolution

have been omitted. In this graph the $L_{R,m,s}$ level for each student has been accumulated by module, hence as shown in Figure 2 those students who did not participate in all the activities proposed, they were considered in the study.

Regarding students' participation, we have monitored participation values (see Figure 3) revealing a decrease of participation level after considering the following information:

– Q: Questionnaire participation.
– F: Total number of post in the forum.
– FP: Participation in the forum.
– P: Peer-to-peer survey participation.

In contrast to the decrease in the participation level, with respect to the evolution of the overall scores in the course, these values are steady along all the modules in the course. The overall scores evolution are shown in Figure 4, which presents the overall score result for each module activity, that is, $L_{R,m,s}$ and $L_{F,m,s}$ without considering each specific student's values and detailing each questions for $L_{R,m,s}$ (i.e. $Q1$, $Q2$ and $Q3$).

We have calculated the correlation coefficient between the values in the point of time 1 to 7 (i.e. each module). The results of the correlation analysis are shown in Figure 5. Pearson's correlation is close to 1 for most of the cases, hence there is a strong relationship between

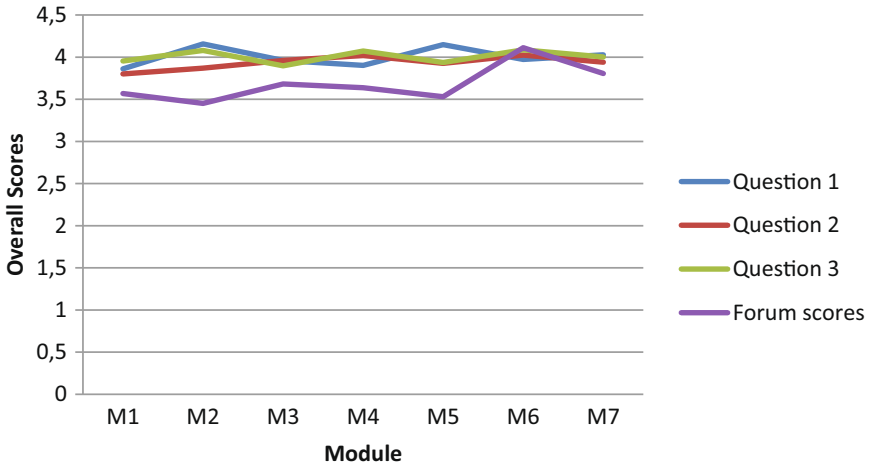

Figure 4 Overall scores in the course

		M1	M2	M3	M4	M5	M6	M7
M1	Pearson Correlation	1,00	**,70**	,64	,54	,59	,54	,63
	Sig. (2-tailed)		,00	,00	,01	,01	,02	,03
	N	40	26	22	20	20	18	12
M2	Pearson Correlation	,70	1,00	**,89**	,81	,86	,81	,69
	Sig. (2-tailed)	,00		,00	,00	,00	,00	,02
	N	26	26	20	18	19	16	11
M3	Pearson Correlation	,64	,89	1,00	**,83**	,76	,80	,79
	Sig. (2-tailed)	,00	,00		,00	,00	,00	,00
	N	22	20	23	19	18	16	12
M4	Pearson Correlation	,54	,81	,83	1,00	**,78**	,76	,80
	Sig. (2-tailed)	,01	,00	,00		,00	,00	,00
	N	20	18	19	20	16	15	11
M5	Pearson Correlation	,59	,86	,76	,78	1,00	**,75**	,90
	Sig. (2-tailed)	,01	,00	,00	,00		,00	,00
	N	20	19	18	16	21	16	11
M6	Pearson Correlation	,54	,81	,80	,76	,75	1,00	**,86**
	Sig. (2-tailed)	,02	,00	,00	,00	,00		,00
	N	18	16	16	15	16	18	12
M7	Pearson Correlation	,63	,69	,79	,80	,90	,86	1,00
	Sig. (2-tailed)	,03	,02	,00	,00	,00	,00	
	N	12	11	12	11	11	12	12

Figure 5 PSPP Pearson coefficient between trustworthiness levels in modules

trustworthiness levels in modules. The observed correlation is positive; consequently, when the trustworthiness level increases in module i, trustworthiness level in module $i + x$ also increases in value. The *sig.* value is less than 0.05, because of this, hence we can conclude that there is a statistically significant correlation between trustworthiness levels. Note that in Figure 5 we have marked those values which correspond to correlation between consecutive module (i.e. $r_{m_i,m_{i+1}}$), in these cases, the coefficient is always more than 0.7.

Finally, in order to compare manual an automatic assessment results, a foremost step is needed. We organized both manual and peer-to-peer activities in a timeline diagram with the aim to compare manual and automatic activities in suitable time references. To this end, we have designed a course plan that permits the comparison process between manual and peer-to-peer assessment. The manual assessment activities are taken as time reference.

Once the time references have been defined, we can compare overall values between manual and automatics method. For instance, Figure 6 shows the dispersion chart between the automatic peer-to-peer activity for the module 1 (i.e. R_1) and the first manual assessment method. It can be seen from the function in Figure 6 that there exist anomalous cases

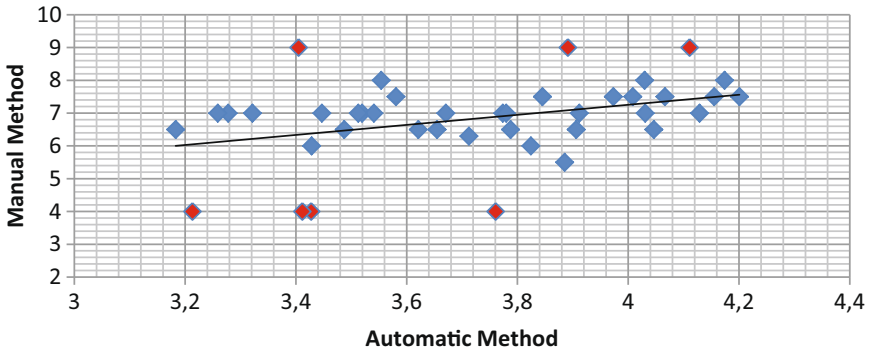

Figure 6 Dispersion chart

detected with respect to the difference between the manual and the automatic value. The rest of the values follow a significant relation between these parameters.

6.6 Findings

In this section we summarize the most relevant findings that emerge from the results and the statistical analysis.

The participation level has experimented a marked decrease along the course, especially at the end of e-assessment activities. We plan to tackle this problem with alternative course schedule with the aim to balance the students' peer-to-peer activities and other students' assignments.

Regarding overall peer-to-peer (i.e. automatic) and continuous (i.e. manual) assessment overall levels, the results reveal a notable difference between the overall range of these values. Figure 6 shows that most of peer-to-peer assessment values are in the range from 3,5 to 4,3 (the e-assessment scale was from 1 to 5) and the continuous assessment, from 1 to 9.

Although the model has to be enhanced and we have to solve the aforementioned problems, the statistical analysis shows significant findings regarding the feasibility of the hybrid evaluation method. The results of the comparisons between manual and automatic assessment indicate (also see Figure 6):

- The mean difference between manual and automatic method is 0,81 (the scale used from 0 to 10).
- The maximum and minimum difference: 0,03 and 2,82.
- The percentage of assessment cases in which the difference between manual and automatic assessment is less than 1 (i.e. 10 % with respect the maximum score) is the 76,92 %.
- If we extend the difference to more than 2 points in the scale, the percentage of assessment cases in this range is the 92,31 %.

The most significant finding is related to anomalous user assessment. From these data, 3 students whose deviation is greater than 20 % were found anomalous and required further investigation for potential cheating in order to validate the authenticity (i.e. identification and integrity) of her learning processes and results.

7 Conclusions and further work

In this paper we have presented an innovative approach for modelling trustworthiness in the context of secure learning assessment in on-line collaborative learning groups. The study shows the need to propose a hybrid assessment model which combines technological security solutions and functional trustworthiness measures. To this end, a holistic security model is designed, implemented and evaluated in a real context of e-Learning. This approach is based on trustworthiness factors, indicators and levels, which allow us to discover how trustworthiness evolves into the learning system.

As ongoing work, we plan to continue the methodology testing and evaluation by deploying e-assessment learning components in additional real on-line courses. Due to further deployments will require large amount of data analysis, we will continue investigating parallel processing methods to manage trustworthiness factors and indicators by improving the MapReduce [14] configuration strategies that would result in improvement of a parallel speed-up, such as customized size of partitions. Moreover, we plan to evaluate and test

trustworthiness predictions methods. With respect to prediction, we would like to improve our approach in order to predict both trustworthiness students' behaviour and evaluation alerts such as anomalous results. To this end, we plan to evaluate neural networks and data mining models by designing a methodological approach to construct a trustworthiness normalized model. In addition, in our future work, we would like to improve our students' public profile model in real on-line courses.

Acknowledgements This research was partly funded by the Spanish Government through the following projects: TIN2011-27076-C03-02 "CO-PRIVACY"; CONSOLIDER INGENIO 2010 CSD2007-0 004 "ARES"; TIN2013-46181-C2-1-R "COMMAS" Computational Models and Methods for Massive Structured Data; and TIN2013-45303-P "ICT-FLAG" Enhancing ICT education through Formative assessment, Learning Analytics and Gamication.

References

1. Bernthal, P.: A survey of trust in the workplace. Executive summary, HR Benchmark Group, Pittsburg, PA (1997)
2. CSO Magazine, US Secret Service, Software Engineering Insistute CERT Program at Carnegie Mellon University, Deloitte: 2011 Cybersecurity Watch Survey. Tech. rep., CSO Magazine (2011)
3. Dai, C., Lin, D., Bertino, E., Kantarcioglu, M.: An approach to evaluate data trustworthiness based on data provenance. In: W. Jonker, M. Petković (eds.) Secure Data Management, vol. 5159, pp. 82–98. Springer, Berlin Heidelberg (2008)
4. Dark, M.J.: Information assurance and security ethics in complex systems: interdisciplinary perspectives. Information Science Reference, Hershey, PA (2011)
5. Demott, J.D., Sotirov, A., Long, J.: Gray Hat Hacking, Third Edition Reviews. 3edn. McGraw-Hill Companies, New York (2011)
6. Eibl, C.J.: Discussion of information security in e-learning. Ph.D. thesis, Universität Siegen. Siegen, Germany (2010). http://dokumentix.ub.uni-siegen.de/opus/volltexte/2010/444/pdf/eibl.pdf
7. Ferencz, S.K., Goldsmith, C.W.: Privacy issues in a virtual learning environment. Cause/Effect, A practitioner's journal about managing and using information resources on college and university campuses, vol. 21, pp. 5–11. Educause (1998). http://net.educause.edu/ir/library/html/cem/cem98/cem9812.html
8. Levy, Y., Ramim, M.: A theoretical approach for biometrics authentication of e-exams. In: Chais Conference on Instructional Technologies Research. The Open University of Israel, Raanana, Israel (2006)
9. Liu, Y., Wu, Y.: A survey on trust and trustworthy e-learning system. In: 2010 International Conference on Web Information Systems and Mining, pp. 118–122. IEEE (2010). doi:10.1109/WISM.2010.62 http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5662295
10. Miguel, J., Caballé, S., Prieto, J.: Providing security to computer-supported collaborative learning systems: An overview. In: Fourth IEEE International Conference on Intelligent Networking and Collaborative Systems (INCOS 2012), pp. 97–104. IEEE Computer Society, Bucharest, Romania (2012). doi:10.1109/iNCoS.2012.60
11. Miguel, J., Caballé, S., Prieto, J.: Security in learning management systems: Designing collaborative learning activities in secure information systems. eLearning Papers. European Comission: elearningeuropa.info. http://elearningeuropa.info/en/article/Security-in-Learning-Management-Systems%3A-Designing-Collaborative-Learning-Activities-in-Secure-Information-Systems?paper=116112 (2012)
12. Miguel, J., Caballé, S., Prieto, J.: Information security in support for mobile collaborative learning. In: The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013), pp. 379–384. IEEE Computer Society, Taichung, Taiwan. doi:10.1109/CISIS.2013.69 (2013)
13. Miguel, J., Caballé, S., Prieto, J.: Providing information security to MOOC: Towards effective student authentication. In: 5-th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2013), pp. 289–292. IEEE Computer Society, Xian, China. doi:10.1109/INCoS.2013.52 (2013)
14. Miguel, J., Caballé, S., Xhafa, F., Prieto, J.: A massive data processing approach for effective trustworthiness in online learning groups. Concurrency and Computation, Practice and Experience (2014)

15. Miguel, J., Caballé, S., Xhafa, F., Prieto, J.: Security in Online assessments: towards an effective trustworthiness approach to support e-learning teams. In: 28th International Conference on Advanced Information Networking and Applications (AINA 2014), pp. 123–130. IEEE Computer Society, Victoria, Canada 2014. doi:10.1109/AINA.2014.106 Best paper of AINA (2014)
16. Miguel, J., Caballé, S., Xhafa, F., Prieto, J., Barolli, L.: A collective intelligence approach for building student's trustworthiness profile in online learning. In: Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2014). GUANGZHOU, P.R. China (2014)
17. Miguel, J., Caballé, S., Xhafa, F., Prieto, J., Barolli, L.: A methodological approach to modelling trustworthiness in online collaborative learning. In: Fourth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2014). Salerno, Italy (2014)
18. Miguel, J., Caballé, S., Xhafa, F., Prieto, J., Barolli, L.: Predicting trustworthiness behavior to enhance security in on-line assessment. In: 6th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2014). Salerno, Italy (2014)
19. Miguel, J., Caballé, S., Xhafa, F., Prieto, J., Barolli, L.: Towards a normalized trustworthiness approach to enhance security in on-line assessment. In: Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2014), pp. 147–154. IEEE Computer Society, Birmingham (2014). doi:10.1109/CISIS.2014.22
20. Mobasher, B., Burke, R., Bhaumik, R., Williams, C.: Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. ACM Trans. Internet Technol. (2007). doi:10.1145/1278366.1278372
21. Ray, I., Chakraborty, S.: A vector model of trust for developing trustworthy systems. In: D. Hutchison, T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M.Y. Vardi, G. Weikum, P.Samarati, P. Ryan, D. Gollmann, R. Molva (eds.) Computer Security - ESORICS 2004, vol. 3193, pp. 260–275. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
22. Schneier, B.: The psychology of security. In: Proceedings of the Cryptology in Africa 1st International Conference on Progress in Cryptology, AFRICACRYPT'08, pp. 50–79. Springer-Verlag, Berlin, Heidelberg (2008)
23. Weippl, E.R.: Security in e-learning. In: H. Bidgoli (ed.) Handbook of information security Vol. 1, Key concepts, infrastructure, standards and protocols., vol. 1, Wiley, Hoboken (2006)
24. Wu, Z., Ou, Y., Liu, Y.: A taxonomy of network and computer attacks based on responses. In: International Conference on Information Technology, Computer Engineering and Management Sciences (ICM), vol. 1, pp. 26 –29 (2011)

2.2 A Massive Data Processing Approach for Effective Trustworthiness in Online Learning Groups

Miguel, J., Caballé, S., Xhafa, F., and Prieto, J. (2015b). A massive data processing approach for effective trustworthiness in online learning groups. *Concurrency and Computation: Practice and Experience (CCPE)*, 27(8):1988–2003. Wiley Online Library. doi:10.1002/cpe.3396. IF: 0.784, Q2: 50/102, Category: COMPUTER SCIENCE, THEORY & METHODS (JCR-2013 SE)

CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. 2015; **27**:1988–2003
Published online 26 September 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cpe.3396

SPECIAL ISSUE PAPER

A massive data processing approach for effective trustworthiness in online learning groups

Jorge Miguel[1,*,†], Santi Caballé[1], Fatos Xhafa[2] and Josep Prieto[1]

[1]*Department of Computer Science, Multimedia, and Telecommunication, Open University of Catalonia, Barcelona, Spain*
[2]*Department of Languages and Informatic Systems, Technical University of Catalonia, Barcelona, Spain*

SUMMARY

This paper proposes a trustworthiness-based approach for the design of secure learning activities in online learning groups. Although computer-supported collaborative learning has been widely adopted in many educational institutions over the last decade, there exist still drawbacks that limit its potential. Among these limitations, we investigate on information security vulnerabilities in learning activities, which may be developed in online collaborative learning contexts. Although security advanced methodologies and technologies are deployed in learning management systems, many security vulnerabilities are still not satisfactorily solved. To overcome these deficiencies, we first propose the guidelines of a holistic security model in online collaborative learning through an effective trustworthiness approach. However, as learners' trustworthiness analysis involves large amount of data generated along learning activities, processing this information is computationally costly, especially if required in real time. As the main contribution of this paper, we eventually propose a parallel processing approach, which can considerably decrease the time of data processing, thus allowing for building relevant trustworthiness models to support learning activities even in real time.

Received 14 July 2014; Accepted 31 August 2014

KEY WORDS: trustworthiness; e-Learning activities; computer-supported collaborative learning; information security; parallel processing; log files; massive data processing; Hadoop; MapReduce

1. INTRODUCTION

Computer-supported collaborative learning (CSCL) has become one of the most influencing educational paradigms [1, 2] widely adopted in many educational institutions over the last two decades. Among these institutions, our real e-Learning context of the Open University of Catalonia[‡] (UOC) develops online education on the basis of collaborative learning activities. This institution is supporting the research work presented in this paper, and its results are considered and included in other UOC's research projects, with the aim of enhancing e-Learning factors, such as assessment cost reduction and students scalability. Although CSCL activities have been incorporated in many online educational settings, there exist still many drawbacks that limit their potential. Among these limitations, collaborative learning services and activities are usually designed and implemented

*Correspondence to: Jorge Miguel, Department of Computer Science, Multimedia, and Telecommunication, Open University of Catalonia, Barcelona, Spain.
[†]E-mail: jmmoneo@uoc.edu

[‡]The Open University of Catalonia is located in Barcelona, Spain. The UOC offers distance education through the Internet since 1994. Currently, about 60,000 students and 3700 lecturers are involved in 8300 online classrooms from about 100 graduate, post-graduate, and doctorate programs in a wide range of academic disciplines. http://www.uoc.edu

without much consideration of security issues. As a result, information security vulnerabilities may interfere in these activities, thus threatening and reducing the effectiveness of the overall collaborative learning process [3, 4].

Information security requirements have been generally considered and developed recently in learning management systems (LMS) [5]. However, to the best of our knowledge, integrated and holistic security models have not been carried out yet. As a result, many security vulnerabilities are still reported in LMSs and remain unsolved [6, 7]. Therefore, innovative security solutions are needed to overcome these limitations and support a secure learning process. To this end, in this paper, we propose a trustworthiness model based on a multifold assessment approach of CSCL activities, which can meet security requirements of online collaborative learning process.

Finally, in order to provide effective and just-in-time trustworthiness information from the LMS, it is required a continuous processing and analysis of group members' interaction data during long-term learning activities, which produces huge amounts of valuable data stored typically in server log files [8, 9]. CSCL activities may demand a great amount of communication processes, collaborative contents, and many types of interactions [1, 2]; if our model aims to analyze how trustworthiness factors are related to these resources, the context of CSCL will be an ideal case study. Because of the large or very large size of data generated daily in online learning activities, the massive data processing is a foremost step in extracting useful information and may require computational capacity beyond that of a single computer [10]. We study the feasibility of a parallel approach for processing large log data files of a real LMS using distributed infrastructures and show how considerable improvements in performance can be achieved via Hadoop MapReduce implementations.

The paper is organized as follows. Section 2 presents the background and context information on security in e-Learning. Section 3 endows our security model with trustworthiness properties on learning activities describing relevant trustworthiness factors and rules that have an effect in the collaborative learning process. Parallel processing paradigms are analyzed in Section 4 to massive data processing and build relevant trustworthiness models. Finally, Section 5 concludes the paper highlighting the main findings and outlining ongoing and future work.

2. BACKGROUND

In this section, we first review main works in the literature on general security in e-Learning, including our previous research. Then, we propose complementary solutions to secure e-Learning beyond technological approaches. To this end, a trustworthiness approach for secure e-Learning is provided.

2.1. Information security in e-Learning

Early research works about information security in e-Learning [11, 12] are focused on confidentiality issues with respect to ensure students' and tutors' privacy requirements. An initial work [13] suggests that the most effective mechanism for dealing with the privacy issues raised in the virtual learning environment should be a task force or committee made up of those who are closely involved. This proposal is quite general, and then, in subsequent works on privacy in e-Learning, some authors have addressed the need for more specific approaches [3]. Further works [4,14] consider other aspects of security in e-Learning. In [14], the author argues that security is mainly an organizational and management issue and improving security is an ongoing process in e-Learning. This is in fact the first proposal in which information security is applied to LMS as a general requirement in e-Learning design and management. The authors in [4] presented how security in e-Learning can be analyzed from a different point of view, namely, by first analyzing threats for e-Learning, and then, recommendations are introduced and discussed in order to cope with detected threats. Finally, more specific security issues in e-Learning have been investigated (e.g., virtual assignments and exams, security monitoring, and authentication and authorization services) in [15–18].

Although the aforementioned literatures discuss on security design in e-Learning from a theoretical point of view, there is still needful to understand attacks in order to discover security design factors and figure out how security services must be classified and designed [19]. In [20], through analyzing

DOI: 10.1002/cpe

existing research in attack classification, a new attack taxonomy is constructed by classifying attacks into dimensions. Nevertheless, because attacks taxonomies might be applied to cover each kind of attack that might occur in LMS, they are not closely related to security design in e-Learning. In order to fill this gap, in [17], we have proposed an alternative approach that associates attacks to security design factors.

There is an increasing interest in understanding security attacks in real-life scenarios. Several reports justify the relevance of security attacks during the last 2 years. In particular, the study presented in [7] revealed that security attacks are a reality for most organizations: 81% of respondents' organizations experienced a security event (i.e., an adverse event that anyhow compromises security). Finally, we can consider specific LMS real software vulnerabilities. Moodle is an open-source LMS that is massively deployed in many schools and universities. In Moodle Security Announcements,[§] 40 serious vulnerabilities have been reported in 2013.

2.2. Previous work on security in e-Learning

In previous research [15–18], we have argued that general security approaches proposed so far do not guarantee that learning processes are developed in a reliable way. Next, we summarize the main research findings on security in e-Learning made so far focused mainly in the following educational contexts: collaborative learning (CSCL), mobile learning (m-Learning), and massive open online courses (MOOCs). These contexts are approached by several design methodologies and security considerations, such as, software modeling languages, risk management, security in LMS, attacks in e-Learning, students' privacy, specific security properties, for example, authentication, and global user authentication services.

In [15], we proposed a new approach named secure collaborative LMS (SCLMS) based on the current developments in the domain of CSCL systems that consider security as a key requirement. As a result, an innovative guideline is proposed to develop secure LMS focusing on the support for CSCL with specific needs, such as interactions between participants, collaborative material management, communication processes, and generation of collaborative results. Following the SCLMS roadmap, in [17], an innovative guideline to develop secure e-Learning systems was presented for m-Learning. In [18], we conducted research to provide information security to the MOOCs [18] and in particular supporting evaluation, grading, and certification as the main challenges in the MOOC arena [21]. The core of this approach is an authentication service defined as a modular PKI-based security model called MOOC Smart Identity Agent (MOOC-SIA) [18], which is a global user authentication model for MOOC platforms.

Considering the previous research experiences, the starting point of this paper is to extend our aforementioned proposals with a new trustworthiness model.

2.3. Trustworthiness and security for e-Learning

In [22], it is discussed that security is both a feeling and a reality. The author points out that the reality of security is mathematical based on the probability of different risks and the effectiveness of different countermeasures. But security is also a feeling, based not on probabilities and mathematical calculations but on your psychological reactions to both risks and countermeasures [22]. This security model eventually concludes that security is a trade-off between the real fact that absolute security does not exist and the need to feel secure. This approach is very relevant in our model because it is based on a hybrid evaluation system in which technological and trustworthiness solutions are combined.

In order to measure trustworthiness and identify what factors are involved in a quantitative study, in [23], it is proposed a data provenance trust model, which assigns trust scores to both data and data providers on the basis of certain factors that may affect trustworthiness. For instance, in our e-Learning context, students and students' resources (e.g., shared documents and posts in a forum) can be modeled following this approach when developing CSCL activities. To this end, in [24], the author designs a

[§]https://moodle.org/mod/forum/view.php?f=996

DOI: 10.1002/cpe

survey to explore interpersonal trust in work groups identifying trust-building behaviors ranked in order of importance. These behaviors can be used as trustworthiness factors, which can measure trust in those activities that students develop. In addition, the authors in [25] consider different aspects of trustworthiness in terms of expressions and classifications of trust characteristics, such as trust asymmetry, time factor, limited transitivity, and reliability.

3. METHODOLOGY

This section shows a methodological approach to build our trustworthiness-based security model for CSCL. First, we built our model by enhancing standard security models with trustworthiness factors and rules, following the considerations made in Section 2. Then, we apply some statistical techniques to the variables involved in our security model with the purpose to measure correlation. Finally, we conclude the section with important issues concerning the management of the large and complex data forming our model, which becomes the main motivation of this research. The next section presents a solution to these issues.

3.1. Trustworthiness for secure e-Learning

Our security model is endowed in this subsection with trustworthiness properties on learning activities and learners themselves. First, we describe relevant trustworthiness factors and rules that have an effect in the collaborative learning process. Then, in order to measure the impact of these factors, we propose several indicators and levels of trustworthiness.

3.1.1. Trustworthiness factors and rules. The relevant trustworthiness factors identified are summarized in the Table I.

In addition, we take into account the following trustworthy rules: (i) asymmetry, where A trust B is not equal to B trust A; (ii) time factor, where trustworthiness is dynamic and may evolve over the time; (iii) limited transitivity, where if A trusts C who trusts B, then A will also trust B, but with the transition goes on, trust will not absolutely be reliable; and (iv) context sensitive, when if context changes, then trust relationship might change too.

However, it is worth mentioning that trustworthiness factors are defined from the perspective of students' behavior; hence, the methods discussed so far provide security improvements but cannot fully meet secure e-Learning activities requirements. Furthermore, neither trustworthiness nor PKI models define or manage the actions to take when the security service detects either anomalous situations or violation of the properties we have defined.

Trustworthiness building and reducing factors are closely related to (see Table I) the following:

Table I. Trustworthiness factors.

	Trustworthiness building factors
	Student (S) working in the group of students (GS) is building trustworthiness when…
1	S communicates honestly, without distorting any information.
2	S shows confidence in GS's abilities.
3	S keeps promises and commitments.
4	S listens to and values what GS say, even though S might not agree.
5	S cooperates with GS and looks for mutual help.
	Trustworthiness reducing factors Student (S) working in the group of students (GS) is reducing trustworthiness when…
1	S acts more concerned about own welfare than anything else.
2	S sends mixed messages so that GS never know where S stands.
3	S avoids taking responsibility.
4	S jumps to conclusions without checking the facts first.
5	S makes excuses or blames others when things do not work out.

DOI: 10.1002/cpe

- Interactions between participants (e.g., TRF2).
- Content management and generation of collaborative results (e.g., TRF1).
- Communication processes (e.g., TBF1).
- Group management tasks (e.g., TRF5).

Every of these issues may be involved in e-Learning, but in CSCL learning experiences, we can find a higher amount of them than in other learning paradigms; hence, we focus our trustworthiness model on CSCL.

3.1.2. Modeling trustworthiness levels and indicators. We introduce now the concept of trustworthiness indicator tw_i (with $i \in I$, where I is the set of trustworthiness indicators) as a measure of trustworthiness factors. Trustworthiness factors have been presented as those behaviors that reduce or build trustworthiness in a collaborative group, and they have been considered in the design of questionnaires. A tw_i is associated with one of the measures defined in each e-assessment instrument (i.e., ratings, questionnaires, and reports). The concept of trustworthiness level Ltw_i is a composition of indicators over trustworthiness rules and characteristics. For instance, we can consider two trustworthiness indicators (tw_a and tw_b). These indicators are different, the first indicator could be a rating in a forum post and the second one a question in a questionnaire, but they measure the same trustworthiness building factor (e.g., TBF-1: communicates honestly). With regard to trustworthiness rules, this indicator may be compared with the group, over the time or considering the context. Trustworthiness indicators can be represented following these expressions:

$$tw_{ar,s}\ a \in \{Q, RP, LGI\}, r \in R, s \in S$$

where Q is the set of responses in questionnaires, RP is the analogous set in reports, LI is the set of LMS indicators for each student (i.e., ratings and the general students' data in the LMS), S is the set of students in the group, and R is the set of rules and characteristics (e.g., time factor). These indicators are described earlier when presenting instruments.

Once indicators have been selected, trustworthiness levels can be expressed as follows:

$$Ltw_i = \sum_{i=1}^{n} \frac{tw_i}{n}, i \in I$$

where I is the set of trustworthiness indicators that are combined in the trustworthiness level Ltw_i.

Trustworthiness levels Ltw_i must be normalized; to this end, we have reviewed the normalization approach defined in [26] with regard to supporting those cases in which particular components need to be emphasized more than the others. Following this approach, we previously need to define the weights vectors:

$$w = (w_1, \ldots, w_i, \ldots w_n), \sum_{i}^{n} w_i = 1$$

where n is the total number of trustworthiness indicators and w_i is the weight assigned to tw_i.

Then, we define trustworthiness normalized levels as

$$Ltw_i^N = \sum_{i=1}^{n} \frac{(tw_i * w_i)}{n}, i \in I$$

Therefore, trustworthiness levels allow us modeling students' trustworthiness as a combination of normalized indicators using research and data gathering instruments.

Regarding groups, this model may also be applied in cases with only one working group; in this scenario, all students would belong to the same group.

DOI: 10.1002/cpe

3.2. Statistical analysis

Following the trustworthiness model presented earlier, we proceed now with inquiring whether the variables involved in the model are related or not. With this purpose, the correlation coefficient may be useful. Some authors have proposed several methods regarding rates of similarity, correlation, or dependence between two variables [27]. Even though the scope of this paper is focused on user-based collaborative filtering and user-to-user similarity, the models and measures of the correlations between two items applied in this context are completely applicable in our scope. More precisely, we propose Pearson correlation coefficient r as a suitable measure devoted to conduct our trustworthiness model. Pearson coefficient applied to a target trustworthiness indicator is defined in the succeeding text:

$$r_{a,b} = \frac{\sum_{i=1}^{n}\left(tw_{a,i} - \overline{tw}_a\right) * \left(tw_{b,i} - \overline{tw}_b\right)}{\sqrt{\sum_{i=1}^{n}\left(tw_{a,i} - \overline{tw}_a\right)^2} * \sqrt{\sum_{i=1}^{n}\left(tw_{b,i} - \overline{tw}_b\right)^2}}$$

where tw_a is the target trustworthiness indicator, tw_b is the second trustworthiness indicator in which tw_a is compared (i.e., similarity, correlation, and anomalous behavior), $\overline{tw}_a$ and $\overline{tw}_b$ are the average of the trustworthiness indicators, and n is the number of student's provided data for tw_a and tw_b indicators.

It is worth mentioning that if both a and b are trustworthiness indicators with several values over the time (e.g., a question that appears in each questionnaire), they must be compared in the same point in time. In other words, it is implicit that $r_{a,b}$ is actually representing r_{a_t,b_t}, where a_t is the trustworthiness indicator in time t.

In addition, this test may be applied to every trustworthiness indicator taking one of them as target indicator. To this end, we define the general Pearson coefficient applied to a target trustworthiness indicator over the whole set of indicators, as follows:

$$r_A = \left(r_{a,1}, \ldots, , r_{a,i}, \ldots, , r_{a,n-1}\right), i \in I, i \neq a$$

where $r_{a,i}$ is the Pearson coefficient applied to a target trustworthiness indicator defined earlier and I is the set of trustworthiness indicators.

Both relation and similarity are represented by $r_{a,b}$ and r_A grouping students' activities and taking the variables at the same time. We are also interested in time factor, and it may be relevant the evolution of trustworthiness indicators throughout the course. To this end, we extend pervious measures, adding time factor variable:

$$r_{a,t,\ tt} = \frac{\sum_{i=1}^{n}\left(tw_{at,i} - \overline{tw}_{at}\right) * \left(tw_{att,i} - \overline{tw}_{att}\right)}{\sqrt{\sum_{i=1}^{n}\left(tw_{at,i} - \overline{tw_a}_t\right)^2} * \sqrt{\sum_{i=1}^{n}\left(tw_{att,i} - \overline{tw_a}_{tt}\right)^2}}$$

where t is the target point in time and tt is the reference point in time (i.e., t is compared against tt); all other variables have already been defined with this case, and they are instanced in two moments in the course.

Similarly, we can calculate $r_{a,t,\ tt}$ for each tt, and then, the following indicator may be used:

$$r_{A,t} = \left(r_{a,1}, \ldots, , r_{a,i}, \ldots, , r_{a,n-1}\right), i \in I, i \neq a$$

The trustworthiness indicators are summarized in Table II.

Finally, trustworthiness indicators may be gathered in a trustworthiness matrix with the aim of representing the whole relationship table for each indicator:

DOI: 10.1002/cpe

Table II. Trustworthiness basic indicators.

	Trustworthiness statistical analysis		
Basic indicators	Description	Group by	Target/reference
$r_{A,b}$	Pearson coefficient applied to a target trustworthiness indicator.	Students	tw_a tw_b
r_a	$r_{a,b}$ over the set of indicators	Indicators	tw_a
$r_{a,t,\ tt}$	Pearson coefficient applied to a tw indicator throughout the course from t to tt.	Time	tw_a t
$r_{A,t}$	$r_{a,t,\ tt}$ over the throughout the course.	Course	tw_a

$$R_{tw} = \begin{bmatrix} 0 & r_{tw_1,tw_2} & \dots & r_{tw_1,tw_n} \\ 0 & 0 & \dots & \dots \\ 0 & 0 & 0 & r_{tw_{n-1},tw_n} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Although the indicators presented are proposed as suitable options for our model, the model will be refined to select those indicators oriented to perform the best similarity and correlation. In addition, this approach is also intended to be a prediction tool, because similarity facts may conduct predictions about the evaluation system and its evolution.

To sum up, the aforementioned indicators, levels, rules, and statistical analysis can become robust instruments to appropriately model trustworthiness in e-Learning groups and eventually extend current security models for CSCL that overcomes many of the limitations reported in Section 2. However, the collection of valuable data and their later statistical analysis to build our security model usually involves the constant processing and analysis of large amounts of ill-formatted information, even in real time, stressing even more the computational cost involved and requiring a high-performance solution to alleviate this cost. For instance, in our real e-Learning context of the UOC, with thousands of online courses and many of them involving e-Learning in work teams, the amount of data collected can be of the scale of 20 GB per day coming from different LMS with different formats, and the information is found with high degree of redundancy, tedious, and ill-formatted as well as incomplete.

The next section presents our parallel data processing approach to overcome this problem in order to make it feasible the construction of security models, such as our trustworthiness-based security model for CSCL presented earlier.

4. PARALLEL PROCESSING APPROACH

In this section, we address the need to alleviate the computational cost of massive processing of the large amounts of data generated during long-term e-Learning activities, with the aim to cope with learner's trustworthiness analysis and the building of trustworthiness models, even in real time. To this end, we propose a parallel approach for massive data processing.

4.1. The problem of processing log files

In previous research [28, 29, 10], we showed that extracting and structuring LMS log data is a prerequisite for later key processes such as the analysis of interactions, assessment of group activity, or the provision of awareness and feedback involved in CSCL. With regard to BSCW, the computational complexity of extracting and structuring BSCW log files is a costly process as the amount of data tends to be very large and needs computational power beyond of a single processor (see Figure 1(A) and also [10,29]). In addition, in [30], we studied the viability of processing very large log data files of a real virtual campus (UOC Virtual Campus) using different distributed

DOI: 10.1002/cpe

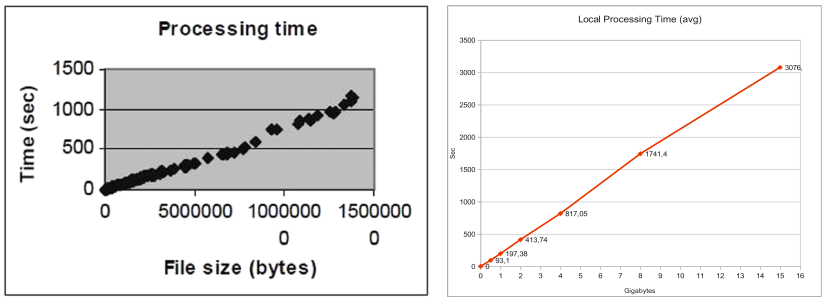

Figure 1. Sequential processing of BSCW log files (A) [10] and local processing of UOC logs (B) [30].

infrastructures to examine the time performance of massive processing of log files. It was also shown the linear execution time of the local processing of UOC log files (Figure 1(B)); hence, the computational cost of sequentially processing large amounts of log data becomes unfeasible.

Therefore, parallel techniques to speed and scale up the structuring and processing of log data are required dealing with log data. In [28] and [30], these models were implemented following the master-slave paradigm and evaluated using Cluster Computing and Planet Lab platforms.

Taking these approaches as starting point, in this paper, we extend our goals in two different directions, which are presented in the next sections: parallelizing the normalization of several LMS logs files (e.g., BSCW and UOC log files) and using MapReduce paradigm [31]. Then, we use Hadoop and Cluster Computing to implement and evaluate the parallelization of massive processing of log data [8].

4.2. UOC Virtual Campus log files

Before presenting our parallel processing implementation details, we first show in this section the different format of BSCW and UOC log files and the problems to process them due to the large size and ill-structure formats of both. To this end, a normalization approach for both types of log data is proposed as an input to our general parallel processing model presented in the next subsections.

4.2.1. BSCW log files. In our real learning context of the Open University of Catalonia, several online courses are provided involving hundreds of undergraduate students and a dozen of tutors in a collaborative learning environment. The complexity of the learning practices entails intensive collaboration activity generating a great amount of group activity information. To implement the collaborative learning activities and capture the group interaction, we use the aforementioned BSCW as a shared workspace system, which enables collaboration over the Web by supporting document upload, group management, and event service among others features. BSCW event service provides awareness information to allow users to coordinate their work [32].

In the BSCW, the events are triggered whenever a user performs an action in a workspace, such as uploading a new document, downloading (i.e., reading) an existing document, and renaming a document. The system records the interaction data into large daily log files and presents the recent events to each user. In addition, users can request immediate e-mail messages whenever an event occurs, and the daily activity reports are sent to them daily and inform them about the events within the last 24 h. The typical format of the BSCW log files is as follows:

```
User:[3434841, '*******']
object:[3452718, 'Presentació A**** S*****']
Type:RateEvent
Time:1078202945.04
Members:[[3448332, '******', 'OyvLkYg2ueStI'], [3449370, '*****', ...., [3425007, 'Aula 5 (*****)'], [3425034, 'Espai per a la Formació de Grups'],
[3425118, 'Espai Presentacions']]
On:[3425118, 'Espai Presentacions']
Touched:[3434844, ':********']
Icon:'/bscw_resources/icons/e_write.gif'
Class:Document
Content:application/octet-stream
```


DOI: 10.1002/cpe

The BSCW log does not follow a standard log format; therefore, parsing these logs format requires a customized development. Moreover, relevant data are omitted. In the example earlier, the student 3434841 is rating the resource 3452718, but we cannot find additional information such as the rate value.

4.2.2. UOC log files. The Web-based virtual campus of the UOC is made up of individual and community virtual areas such as mailbox, agenda, classrooms, library, and secretary's office. Students and other users (lecturers, tutors, administrative staff, etc.) continuously browse these areas where they request for services to satisfy their particular needs and interests. For instance, students make strong use of e-mail service so as to communicate with other students and lecturers as part of their learning process. All users' requests are chiefly processed by a collection of Apache[¶] web servers as well as database servers and other secondary applications, all of which provide service to the whole community and thus satisfy a great deal of users' requests. For load balance purposes, all HTTP traffic is smartly distributed among the different Apache web servers available. Each Web server stores in a log file all users' requests received in this specific server and the information generated from processing the requests. Once a day (namely, at 01:00 AM), all web servers in a daily rotation merge their logs producing a single very large log file containing the whole user interaction with the campus performed in the last 24 h. A typical daily log file size may be up to 20 GB. This great amount of information is first preprocessed using filtering techniques in order to remove a lot of futile, irrelevant information (e.g., information coming from automatic control processes and the uploading of graphical and format elements). However, after this preprocessing, about 2.0 GB of potentially useful information corresponding to 5,000,000 of log entries in average still remains [30].

The log files storing the whole activity of the UOC Virtual Campus follow the Apache log system. A typical configuration for the Apache log system is the Common Log Format [33]; a standard configuration for this log system is as follows:

$$LogFormat"\%h\%l\%u\%t\ \ "\%r\ "\%>s\%b"common$$

where *h* is the IP address of the client or remote host, *l* indicates unavailable requested information, *u* is the user id, *t* is the time that the server finished processing the request, *r* is the request line, *s* is the status code, and *b* is the size of the object returned.

In UOC Virtual Campus, log file records are managed following a variation of the Common Log Format known as Combined Log Format [33], with two additional fields:

$$LogFormat"\%h\%l\%u\%t\ "\%r\ "\%>s\%b\ \ "\%\{Referer\}i\ "\ "\%\{User\text{-}agent\}i\ ""combined$$

where *Referer* field shows the site that the client reports having been referred from and *User-agent* field identifies information that the client browser reports about itself.

As an example, the following is a record that is part of a real log of the UOC Virtual Campus (IP address has been anonymized):

[15/Mar/2012:00:26:40 + 0100] xxx.xxx.xxx.xxx "POST /WebMail/listMails.do?mensajeConfirmacion = El%20missatge%20s´ha%20desplaçat%20a%20la%20carpeta%20Rectorat HTTP/1.1" 200 "http://cv.uoc.edu/WebMail/readMail.do" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 6.1; Trident/5.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; OfficeLiveConnector.1.3; OfficeLivePatch.0.0; InfoPath.2; BRI/2)" 8857 20A

This record example illustrates that the user ID parameter described in Apache Combined Log Format is not available in this line. Moreover, both unavailable requested information (*l*) and the size of the object (*b*) do not meet the standard arrangement. Although user identifications are not stored in log files, the system maintains a session ID, this value is a user session key (a 128-character string long) included as a parameter in the request.

[¶]http://httpd.apache.org/

DOI: 10.1002/cpe

At this point, we highlight certain problems arisen by dealing with these log files:

- We can identify uniquely neither the user nor the record.
- Each explicit user request generates at least an entry in the log file, but it usually generates additional requests; for instance, in order to compose a user Web interface, each component (i.e., image and style sheets) will be loaded using GET operations. This information is not relevant, and these records unnecessarily increase both the storage space and processing effort.
- Additional parameters introduced in Combined Format (Referer and User-agent) may be useful for audit purposes, but in our context, this values introduce a high degree of redundancy.

Because these problems must be solved, we propose several actions. In the case of the user identification, this limitation cannot be completely solved because this information is unavailable, but regarding records identification, it is possible to combine several fields as a record key. The parameters selected to identify a record are

$$Record_{Key} = \{IP; Time; Session\}$$

Redundant and unnecessary information must be parsed and ignored. To this end, these actions have been implemented in the Java class *Action*, which is described in the following subsection following a record taxonomy devoted to clean unusable data. Moreover, regarding storage space, we next also propose which the most efficient way to store record data is.

4.2.3. Log file normalization. Log data normalization or unification is gaining attention from the autonomic computing community [34] as a way to transform proprietary and heterogeneous formatted log data to a standard log data format.

In [28], the task of structuring event log data can be defined as the processes that provide structure to the semi-structured textual event log data and persist the resulting data structure for the later processing by analysis tools. Real e-Learning scenarios usually are formed by several LMS. Therefore, the input of the process is a set of LMS logs files generated by each source. As shown earlier, every log file, such as BSCW and UOC, has its own format showing strong differences in the formatting styles (e.g., in UOC Virtual Campus, a log record is a text line in the text file whilst in BSCW each line represents an attribute value). Moreover, we cannot consider either unifying or normalizing those logs generated by the same Web Server (e.g., both Moodle and UOC Virtual Campus use Apache Web Server, but they log different information stored in different format); hence, a preliminary process is needed in order to normalize these sources following a unified format. To this end, we propose the following tuple:

$$L = (u, t, a, [v]^*)$$

which represents user u performing an action a, which occurs in time t. A list of values $[v]^*$ is associated to the action. An example of a $(a, [v]^*)$ instance could be

$$(create_{document}, document.txt, 1024KB)$$

where first action-value is the filename of the document and the second is the size of the document.

Once we have normalized the log files, the resulting data structure persists for later data processing and analysis [28]. Next, we proceed with a log data processing approach.

4.3. Parallel processing approach

The parallel implementation in the distributed infrastructures that we propose in this subsection follows the MapReduce paradigm [31]. Therefore, we introduce first our MapReduce model on the normalization of different LMS log files, namely, BSCW and UOC, described in the previous section. The results obtained will conduct our parallel implementation approach based on Hadoop and Cluster Computing presented in the next sections.

DOI: 10.1002/cpe

4.3.1. MapReduce paradigm. We can assume that each log file type is a semi-structured text file with record-oriented structure and the input data set is made up of a large number of files storing log information (e.g., each LMS and log per day). The input may be represented as

$$I = \left\{Log_l^i\right\}, l \in L, i \in I$$

where L is the set of LMS and I is the set of log files in an LMS.

The MapReduce paradigm works by splitting the processing into two stages, the map phase and the reduce phase, and each phase has key-value pairs as input and output. Therefore, we define the tasks in the map phase and those processed in the reduce phase, selecting the input and output keys for each phase. In this paradigm, the output from the map function could be processed by the framework before being sent to the reduce function.

The Map phase takes as input a record stored into a log file in I; the key of this record is the offset in the file. When the map function receives the record, it will be processed following the normalization process, which was presented earlier, and this output will be the input for the reduce function. At this point, we can decide among several alternatives dealing with reduce function. In order to only store normalized data, the reduce task does not perform additional work and store the output of map function in the distributed file system. In addition, the reduce function may be used to compute a relevant component as presented in the previous section. In that case, one of the keys is the student, and the reduce function calculates the result of the parameter selected (e.g., number of documents created by the student, total session time, and sum of ratings).

4.3.2. Hadoop. The abstract model proposed in the aforementioned section supporting the MapReduce paradigm will be implemented in the parallel platform of Apache Hadoop.[‖] In [31], MapReduce is presented as a model oriented to further implementations in Hadoop; hence, we take this work as main reference in order to design our normalization LMS log files MapReduce framework.

Hadoop MapReduce job is defined as a unit of work that the client wants to be performed [31] consisting in the input data, the MapReduce program, and configuration information. Then, Hadoop runs the job by dividing it into tasks of two types: map tasks and reduce tasks. There are also two types of nodes: job tracker, which coordinates the paralleling process, and several workers that perform the target work. Hadoop divides the input to a MapReduce job into fixed-size pieces and creates one map task for each split, which runs the map function for each record in the split. It is important to note that the number of reduce tasks is not ruled by the size of the input.

The implementation of map and reduce function is based on these previous works [28,30], which deal with BSCW and UOC Virtual Campus log formats. Once the logs are computed by the event extractor functions, the output is normalized following the model presented.

4.3.3. Record taxonomy and implementation. The implementation of our parallel approach is in Java, which is compatible with Hadoop. Although certain implementation details are omitted, in this section, we present the main services developed. These services are based on a study of the types of records registered in UOC Virtual Campus log files.

The core of the service is implemented in the Java class *Action*, which offers the main methods to process a record (i.e., a log file line). An action object represents something that has occurred in the Virtual Campus as described in UOC Virtual Campus log files. The main services and functions offered by the class Action are a set of get methods (e.g., get_date(), get_ip(), and get_junk()) intended to parse the log line, and the following classification summarizes the output records, which can be managed using these methods:

- Record (R). Logs file line.
- Invalid record (IR). An IR is a record that does not have a valid key. As previously stated, a valid key is a tuple with these components: session, IP, and time.
- Valid record (VR). In contrast, a VR contains each necessary field to form a valid record key.

[‖]http://hadoop.apache.org

DOI: 10.1002/cpe

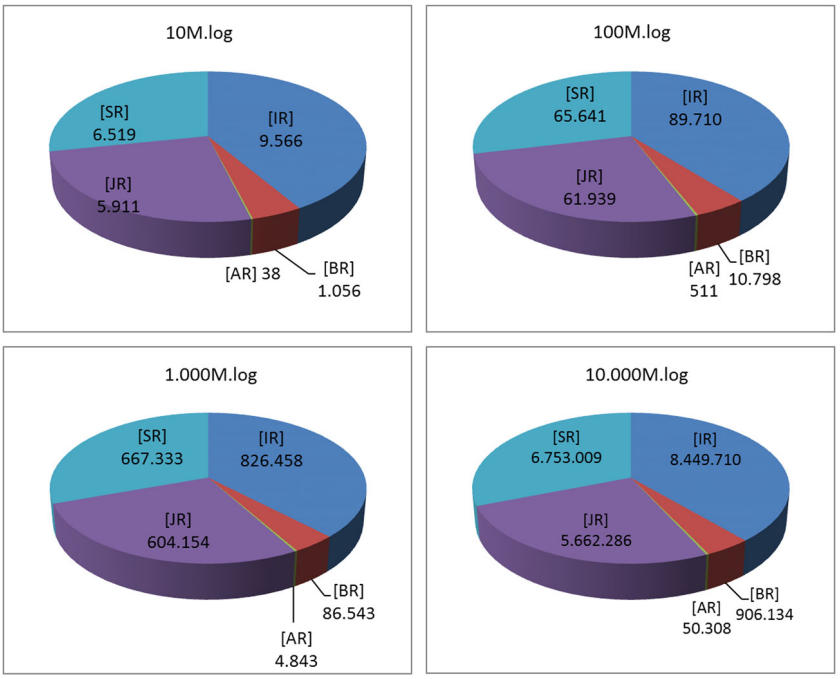

Figure 2. Types of records benchmarks.

- Request record (RR). This type of record is a VR that has requests, and the server generates a 200 return code value. The set of CR includes Junk, Analysis, and System records (which are defined in the succeeding text).
- Short record (SR). If a VR does not reach conditions of RR (i.e., request and 200 code).
- Junk record (JR). We define a JR as an RR on which we can ensure that does not contain relevant data. For instance, an image file requested by the client when creating a user Web interface. In other words, a JR is valid, but it does not contain useful data.
- Analysis record (AR). Over the set of CR and those that are not JR, we select such records, which are relevant to a specific analysis. As we will present in the succeeding text, we select 11 representative actions that a student may perform in the UOC Virtual Campus (e.g., an action may be a student accessing to a classroom environment).
- System record (SR). Because the set of AR is selected for a specific case study, there exists a certain amount of request records that are not considered in the analysis. We name this type of records system records.

4.3.4. Type of records. Of particular relevance is the amount of records computed of each type described earlier. These results will determine the best approach, sequential or parallel, to design the processing log model.

We run four types of records benchmarks for 10, 100, 1.000, and 10.000 MB log files. Results of these tests are shown in the succeeding text (Figure 2):

As can be observed in Figure 2, the amount of AR over the total data set is very low. Therefore, we need a preprocessing phase in order to extract useful information from logs files. Moreover, the average of each type of record is similar over the four tests; the size of the file does not generate different averages. Even when extending the study to more than the 11 selected actions in AR, the amount of BR is also very low.

4.3.5. Results for analysis records. We select those records that are relevant to our specific analysis associated to the actions performed in the LMS,** which must be analyzed. Table III

**Although we focus on students' e-Learning and behavior actions, additional technological information, such as students' device or IP control, could be also included in the study.

DOI: 10.1002/cpe

Table III. Trustworthiness basic indicators.

		Benchmark (xMB)										
Name	Description	1	10	50	250	100	500	1000	2000	4000	8000	10,000
Classroom	Access to a classroom environment	2	21	118	229	603	1112	2222	4429	8861	17,624	22,388
Login	Login LMS session	0	4	18	47	106	217	461	904	1717	3623	4617
Logout	Logout LMS session	1	2	3	8	17	23	67	155	325	747	954
File	Download a file	0	7	57	119	253	524	934	1982	4008	7841	9958
Mail	Load the e-mail service	0	0	1	3	11	37	118	232	466	835	973
Community	Community campus	0	0	3	9	22	55	123	263	536	1125	1390
Services	General services	0	0	1	7	19	38	66	126	273	521	642
Secretary	Secretary's office service	0	2	13	33	69	127	295	648	1283	2563	3081
Profile	Load an user profile	0	1	7	12	22	40	69	127	235	472	592
News	UOC news service	0	1	7	14	28	41	78	184	431	901	1087
Help	Help Desk	0	0	1	2	6	11	20	49	127	264	316

LMS, learning management systems; UOC, Open University of Catalonia.

shows the name of each action, a short description, and the number of user actions computed in each input log file.

4.4. Hadoop processing logs implementation

In this section, we present the most significant aspects of deploying and implementing a MapReduce paradigm intended to manage log data as described in this paper.

4.4.1. MapReduce Java implementation. MapReduce Java implementations are completely based on the class *Action*, which was early developed to test sequential results with regard to time and records types benchmarks. We have developed two separated Java applications, which are presented in this section.

UOCLogDriverClean program normalizes UOC logs files cleaning unnecessary data. Only records in the AR set are considered as outcome, and the other record sets are ignored. In order to improve computational performance, the algorithm progressively inspects each condition in a well-arranged way, that is, first the most restrictive and general condition (e.g. *has_session*) and finally the most specific one (e.g., *has_opa*). The mapper receives as parameters a pair (key, value), where the key is automatically generated by Hadoop and the value is a line of a log file. The output is a different pair where the key is the record ID and the value is the request. It is important to note that, in this case, we do not use reduce function because we are not running reduce tasks (i.e., grouping and computing).

UOCLogDriverCountOp is the second application developed, and it has both map and reduce functions. In this case, our goal is computing aggregate data by computing the sum of each action type (the same outcome as in sequential implementation). The *UOCLogDriverCountOp* map code is similar to *UOCLogDriverClean* implementation; however, output key value for the type of value has been denied as integer to compute each instance. When map function has generated each key-value pair, the output is combined over the value key and processed by the reduce function.

Finally, collected data are stored in the output directory, which is defined when the job is executed.

DOI: 10.1002/cpe

Table IV. Comparative MapReduce results.

	Log file size											
Nodes	1	10	50	250	100	500	1000	2000	4000	8000	10000	Speed up (%)
0	0	0	2	2	9	19	35	75	141	288	353	
2	14	14	15	14	15	29	44	77	141	280	339	4%
4	15	15	15	14	15	20	27	44	74	134	170	52%
6	14	14	16	15	15	15	25	38	64	117	151	57%
8	16	14	15	16	16	16	21	33	44	83	102	71%
10	14	22	15	17	16	21	16	33	37	72	83	76%

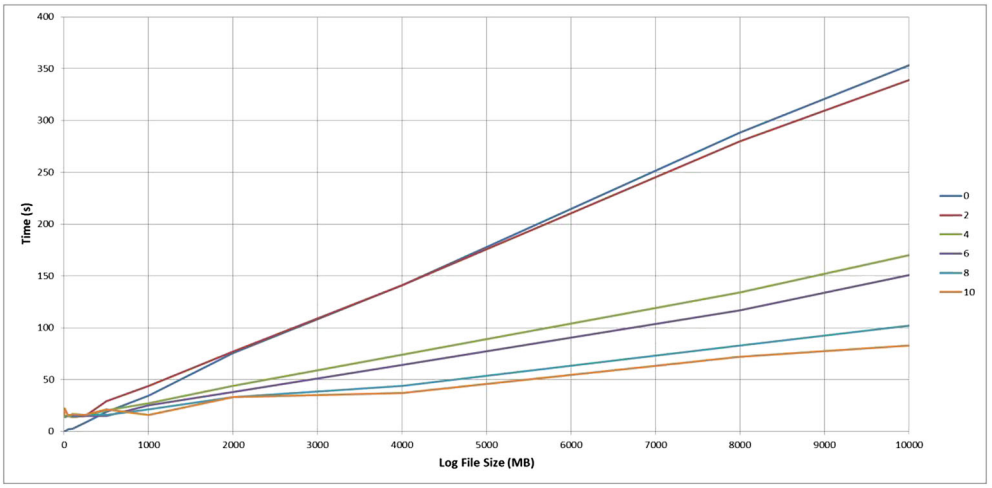

Figure 3. Comparative MapReduce results.

4.5. Analysis of the results

Once the MapReduce applications have been developed, before running the jobs in parallel processing, network and distributed file systems are needed. Hadoop Distributed File System (HDFS) supports large data sets across multiple hosts to achieve parallel processing. HDFS is a block-structured file system based on splitting input data into small blocks of fixed size, which are delivered to each node in the cluster. We use HDFS as Hadoop MapReduce storage solution; therefore, some file system configuration tasks are needed, such as create user home file and define suitable owner, create MapReduce jobs input directory, upload log files, and retrieve results actions.

In Table IV and Figure 3, we can see comparative results of the battery of tests with multiple Hadoop nodes (i.e., 2, 4, 6, 8, and 10 workers) in RDlab[††] cluster. Note that 0 node shows results of local sequential processing benchmark. Furthermore, we have carried out additional file system integration processes by running Hadoop jobs over the open-source Lustre[‡‡] file system, which is deployed in the RDlab.

From this experimental study, we can see that the results do not grow linearly anymore. We can also see that by using a distributed MapReduce Hadoop infrastructure, a considerable speed up is achieved in processing large log file data as shown in Table IV, last column, (i.e., more than 50% for infrastructures with more than four nodes and more than 75% for 10 nodes). Regarding log file size, for too small values, the overhead introduced by the MapReduce framework, when sending the parts to the nodes and combining output data, is noticeable, and the framework control tasks spends too much time managing and distributing data. On the other side, values of the task size close to

[††]http://rdlab.lsi.upc.edu
[‡‡]http://lustre.opensfs.org/

DOI: 10.1002/cpe

3000 MB considerably diminish this amount of time in comparison with the total processing time. Moreover, Reduce tasks spend too much time when the number of nodes is low.

5. CONCLUSIONS AND FURTHER WORK

In this paper, we first motivated the need to improve information security in e-Learning and in particular in CSCL activities. Then, we proposed a methodological approach to build a security model for CSCL activities with the aim to enhance standard technological security solutions with trustworthiness factors and rules. As a result, the guidelines of a holistic security model in online collaborative learning through an effective trustworthiness approach were first proposed. However, as learners' trustworthiness analysis involves dealing with large amount of data generated along learning activities, processing this information is computationally costly, especially if required in real time. To this end, and as a main contribution of this paper, we proposed a parallel processing approach that can considerably decrease the time of data processing, thus allowing for building relevant trustworthiness models to support learning activities even in real time.

The implementation of our parallel approach faced two important challenges: handle several formats of logs files coming from different LMS and the large size of these log files. We showed how to normalize different log file structures as an input for the MapReduce paradigm to manage huge amounts of log data in order to extract the trustworthiness information defined in our model.

Finally, we used distributed infrastructure, such as Hadoop and Cluster Computing, to implement and evaluate our parallelization approach for massive processing of log data. The experimental results showed the feasibility of coping with the problem of structuring and processing ill-formatted, heterogeneous, large log files to extract information on trustworthiness indicators and levels from learning groups and ultimately fill a global framework devoted to improve information security in e-Learning in real time. We eventually conclude that it is viable to enhance security in CSCL activities by our trustworthiness model, though taking on the overhead caused by the use of distributed infrastructure for massive data processing.

As ongoing work, we plan to improve the MapReduce configuration strategies that would result in improvement of a parallel speedup, such as customized size of partitions. Furthermore, we are investigating normalized trustworthiness improvements to extend the model presented in this paper.

ACKNOWLEDGEMENTS

This research was partly funded by the Spanish Government through the following projects: TIN2011-27076-C03-02 'CO-PRIVACY', CONSOLIDER INGENIO 2010 CSD2007-0 004 'ARES', TIN2013-46181-C2-1-R 'COMMAS' Computational Models and Methods for Massive Structured Data, and TIN2013-45303-P 'ICT-FLAG' Enhancing ICT Education through Formative Assessment, Learning Analytics and Gamication.

REFERENCES

1. Koschmann T. Paradigm shifts and instructional technology. In *CSCL: Theory and Practice of an Emerging Paradigm*, Koschmann T (ed). Mahwah, New Jersey: Lawrence Erlbaum Associates, 1996; 1–23.
2. Dillenbourg P. What do you mean by collaborative learning? In *Collaborative-Learning: Cognitive and Computational Approaches*, Dillenbourg P (ed.). Elsevier Science: Oxford, UK, 1999; 1–19.
3. Weippl ER. *Security in e-Learning*. Springer: New York, NY, 2005.
4. Eibl CJ. Discussion of information security in e-learning. Universität Siegen: Siegen, Germany, 2010.
5. Kambourakis G, Kontoni D-PN, Rouskas A, Gritzalis S. A PKI approach for deploying modern secure distributed e-learning and m-learning environments. *Computers & Education* 2007; **48**(1):1–16.
6. Trustwave. Trustwave 2012 Global Security Report. Trustwave, 2012.
7. CSO Magazine, US Secret Service, Software Engineering Institute CERT Program at Carnegie Mellon University, and Deloitte. 2011 Cybersecurity Watch Survey. CSO Magazine, 2011.
8. Narkhede S, Baraskar T. HMR log analyzer: analyze Web application logs over Hadoop MapReduce. *International Journal of UbiComp* 2013; **4**(3):41–51.
9. Ciesielski V, Lalani A. Data mining of web access logs from an academic web site. In *Design and Application of Hybrid Intelligent Systems*, Abraham A, Köppen M, Franke K (eds). IOS Press: Amsterdam, The Netherlands, The Netherlands, 2003; 1034–1043.

DOI: 10.1002/cpe

10. Caballé S, Paniagua C, Xhafa F, Daradoumis T. A grid-aware implementation for providing effective feedback to online learning groups. In *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops*, Meersman R, Tari Z, Herrero P (eds), vol. **3762**. Springer Berlin Heidelberg: Agia Napa, Cyprus, 2005; 274–283.
11. Borcea K, Donker H, Franz E, Pfitzmann A, Wahrig H. Privacy-aware eLearning: why and how. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA) 2005*, Chesapeake, Virginia, 2005; 1466–1472.
12. Eibl CJ. Privacy and confidentiality in e-learning systems. In *Proceedings of the Fourth International Conference on Internet and Web Applications and Services*, 2009 - ICIW 09, 2009; 638–642.
13. Ferencz SK, Goldsmith CW. Privacy issues in a virtual learning environment. *Cause/Effect, A practitioner's journal about managing and using information resources on college and university campuses* 1998; **21**:5–11.
14. Weippl ER. Security in e-learning. In *Handbook of Information Security Vol. 1, Key Concepts, Infrastructure, Standards and Protocols*, vol. 1, 3 vols. John Wiley & Sons, Inc.: Hoboken, NJ, 2006; 279–294.
15. Miguel J, Caballé S, Prieto J. Providing security to computer-supported collaborative learning systems: an overview. In *Proceedings of the Fourth IEEE International Conference on Intelligent Networking and Collaborative Systems (INCOS 2012)*, Bucharest, Romania, 2012; 97–104.
16. Miguel J, Caballé S, Prieto J. Security in learning management systems: designing collaborative learning activities in secure information systems. *eLearning Papers. European Comission: elearningeuropa.info*, 2012.
17. Miguel J, Caballé S, Prieto J. Information security in support for mobile collaborative learning. In *Proceedings of the 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013)*, Taichung, Taiwan, 2013; 379–384.
18. Miguel J, Caballé S, Prieto J. Providing information security to MOOC: towards effective student authentication. In *Proceedings of the 5-th International Conference on Intelligent Networking and Collaborative Systems INCoS-2013*, *Xian, China, 2013; 289–292.*
19. Demott JD, Sotirov A, Long J. *Gray Hat Hacking, Third Edition Reviews* (3rd edn). McGraw-Hill Companies: New York, 2011.
20. Wu Z, Ou Y, Liu Y. A taxonomy of network and computer attacks based on responses. In *Proceedings of the 2011 International Conference of Information Technology, Computer Engineering and Management Sciences* 2011; **1**:26–29.
21. Pappano L. The year of the MOOC. *The New York Times*, 2012.
22. Schneier B. The psychology of security. In *Proceedings of the Cryptology in Africa 1st international conference on Progress in cryptology*, Berlin, Heidelberg, 2008; 50–79.
23. Dai C, Lin D, Bertino E, Kantarcioglu M. An approach to evaluate data trustworthiness based on data provenance. In *Secure Data Management*, Jonker W, Petković M (eds), vol. **5159**. Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; 82–98.
24. Bernthal P. *A Survey of Trust in the Workplace*. HR Benchmark Group, Pittsburg, PA, Executive Summary, 1997.
25. Abdul-Rahman A, Hailes S. Using recommendations for managing trust in distributed systems. In *Proceedings of the IEEE Intl. Conference on Communication, Malaysia*, 1997.
26. Ray I, Chakraborty S. A vector model of trust for developing trustworthy systems. In *Computer Security – ESORICS 2004*, Samarati P, Ryan P, Gollmann D, Molva R (eds), vol. **3193**. Springer Berlin Heidelberg: Berlin, Heidelberg, 2004; 260–275.
27. Mobasher B, Burke R, Bhaumik R, Williams C. Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness. *ACM Transactions on Internet Technology* 2007; **7**(4).
28. Xhafa F, Paniagua C, Barolli L, Caballe S. A parallel grid-based implementation for real-time processing of event log data of collaborative applications. *International Journal of Web and Grid Services* 2010; **6**(2):124–140.
29. Caballe S, Xhafa F, Daradoumis T. A grid approach to efficiently embed information and knowledge about group activity into collaborative learning applications. In *The Learning Grid Handbook*, Vol. **2**. IOS Press: Amsterdam, The Netherlands, 2008; 173–197.
30. Caballé S, Xhafa F. Distributed-based massive processing of activity logs for efficient user modeling in a virtual campus. *Cluster Computing* 2013; **16**(6):829–844.
31. White T. *Hadoop: The Definitive Guide* (Third edn). O'Reilly: Beijing, 2012.
32. Appelt W. What groupware functionality do users really use? Analysis of the usage of the BSCW system; 337–341.
33. The Apache Software Foundation. Apache HTTP Server Version 2.2 Documentation, 2014. [Online]. Available: http://httpd.apache.org/docs/2.2/.
34. Salfner F, Tschirpke S, Malek M. Comprehensive logfiles for autonomic systems. In *Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International*, 2004; 211.

DOI: 10.1002/cpe

2.3 A Methodological Approach for Trustworthiness Assessment and Prediction in Mobile Online Collaborative Learning

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2015c). A methodological approach for trustworthiness assessment and prediction in mobile online collaborative learning. *Computer Standards & Interfaces (CSI)*. Springer. doi:10.1016/j.csi.2015.04.008. IF: 1.177, Q2: 38/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE)

A Methodological Approach for Trustworthiness Assessment and Prediction in Mobile Online Collaborative Learning

Jorge Miguel[a,*], Santi Caballé[a], Fatos Xhafa[a], Josep Prieto[a], Leonard Barolli[b]

[a] *Department of Computer Science, Multimedia, and Telecommunication, Open University of Catalonia, Barcelona, Spain*
[b] *Fukuoka Institute of Technology, Department of Information and Communication Engineering, Fukuoka, Japan*

Abstract

Trustworthiness and technological security solutions are closely related to online collaborative learning and they can be combined with the aim of reaching information security requirements for e-Learning participants and designers. Moreover, mobile collaborative learning is an emerging educational model devoted to providing the learner with the ability to assimilate learning any time and anywhere. In this paper, we justify the need of trustworthiness models as a functional requirement devoted to improving information security. To this end, we propose a methodological approach to modelling trustworthiness in online collaborative learning. Our proposal sets out to build a theoretical approach with the aim to provide e-Learning designers and managers with guidelines for incorporating security into mobile online collaborative activities through trustworthiness assessment and prediction.

Keywords: information security, trustworthiness, assessment, prediction, online collaborative learning, mobile learning

1. Introduction

Over the last decade, Computer Supported Collaborative Learning (CSCL) has become one of the most influencing paradigms devoted to improving e-Learning [1]. Similarity, mobile learning is an emerging educational model devoted to providing the learner with the ability to assimilate learning any time and anywhere [2]. Mobile learning provides ubiquity and pervasiveness, which have become essential requirements to support learning and allow all learning community members from a variety of locations to cooperate with each other by means of a large variety of technological equipment [3]. While there has been an explosion of mobile devices and applications in the marketplace to gain access to e-Learning systems and collaborative learning processes, the development of mobile supported collaborative learning guided by technological security as a key and transverse factor has been, to the best of our knowledge, little investigated [4]. However, Information and Communication Technologies (ICT) have been widely adopted and exploited in most of educational institutions in order to support e-Learning through different learning methodologies, ICT solutions and design paradigms. In this context, e-Learning designers, managers, tutors, and students are increasingly demanding new requirements. Among these requirements, information security is a significant factor involved in e-Learning processes. However, according to [5, 6], e-Learning services are usually designed and implemented without much consideration of security aspects. This finding has been usually tackled with ICT security solutions, but as stated in [7], the problems encountered in ensuring modern computing systems, cannot be solved with ICT alone. In contrast, current advanced ICT security solutions are feasible in many e-Learning scenarios though assessment processes in CSCL involve specific non-technological components. Indeed, online assessment activities (e-assessment) usually have specific issues, such as student's grades or course certification, that e-Learning designers have to consider when they manage security requirements. In this context, even most advanced and comprehensive technological security solutions cannot cope with the whole domain of e-Learning vulnerabilities.

*Corresponding author
Email address: jmmoneo@uoc.edu (Jorge Miguel)

An e-Learning activity is a general concept that can involve very different cases, actors, processes, requirements, and learning objectives in the complex context of e-Learning [8]. To conduct our research we focus on specific online collaborative activities, namely, online assessment (e-assessment). In [9], the authors report that the e-assessment process offers enormous opportunities to enhance the student's learning experience, such as delivering on-demand tests, providing electronic assessment, and immediate feedback on tests. In this context, e-assessment is considered an e-exam with most common characteristics of virtual exams, and is typically employed to deliver formative tests to the students. An e-assessment activity is an e-exam with most common characteristics of virtual exams. Moreover, in [10] it is discussed how unethical conduct during e-learning exam-taking may occur and an approach that suggests practical solutions based on technological and biometrics user authentication is proposed.

In our real context of online higher education, we mainly consider peer-to-peer assessment processes and online collaborative activities, which will form e-assessment components. In this context, we propose security technological solutions extended with a functional trustworthiness approach [11, 12, 13] by proposing a hybrid assessment method based on trustworthiness models. From these previous works, in this paper, we endow trustworthiness models for security in e-Learning with a trustworthiness methodology. This approach is devoted to improving security in CSCL by building a trustworthiness methodology to offer guidelines for designing as well as managing security in online collaborative activities, through trustworthiness assessment and prediction. To this end, we propose a trustworthiness methodology with the aim of managing and predicting reliable assessment processes in e-assessment. As a result, by predicting collaborative e-assessment results, e-Learning designers will be able to manage assessment process with additional information generated by automatic prediction models.

This paper is structured as follows. In Section 2 we review the main works in the literature on mobile collaborative learning and security in CSCL, how trustworthiness assessment and prediction are related to security, and trustworthiness methodologies. In Section 3, we describe the theoretical features, phases, data, and processes of our methodological approach. In order to validate and support the application of the methodology, in Section 4 we concrete the most significant aspects in terms of specific methods through their application in real online courses. Moreover, in Section 5 we present and evaluate a neural network approach for peer-to-peer e-assessment prediction. Finally, conclusions and further work are presented in Section 6.

2. Background

In this section, we review the main works in the literature on mobile collaborative learning and security in collaborative learning, how trustworthiness assessment and prediction are related to security, and trustworthiness existing methodologies.

2.1. Security in Online Collaborative Learning

According to [1], Computer Supported Collaborative Learning has become one of the most influencing educational paradigms devoted to improving e-Learning. Some authors argued that information security has to be considered with the aim of ensuring information managed in CSCL. In addition, several technological solutions were proposed [5, 6]. These security solutions, based on technological approaches, tackle the security in e-Learning problem with specific methods and techniques that deal with particular security issues, but these models do not offer an overall security solution [4, 14]. One of the key strategies in information security is that security drawbacks cannot be solved with technology solutions alone [7]. Even most advances security ICT solutions have drawbacks that impede the development of complete security frameworks.

Finally, some authors argue that we need to understand attacks in order to discover relevant security design factors [15]. Real-life security attacks and vulnerabilities are presented in many security reports, which justify the relevance of security attacks over the last years [16, 17].

2.2. Mobile Collaborative Learning

Mobile learning has lately emerged with the increasing use of mobile technology in education. According to [2] and [3] the needs of educational organizations are increasingly related to modern online learning environments which must provide advanced capability for the distribution of learning activities and the necessary functionalities and learning resources to all participants, regardless of where these participants and resources are located, and whether

this location is static or dynamic. The aim of newest learning environments is to enable the learning experience in open, dynamic, large-scale, and heterogeneous environments.

Although, from a general point of view, mobile learning can be considered as any time and anywhere learning experiences, [18] shows how we can consider multiple definitions of m-Learning. Moreover, because of the complexity and multidisciplinary factors of Mobile Computer Supported Collaborative Learning (MCSCL) paradigm, in [3] a three-dimensional approach has been provided to understand and unify the rather dispersion currently existing in advanced learning practices and pedagogical goals from the era of MCSCL. This approach considers the context of MCSCL from a multiple dimensional perspective: pedagogical, technological and evaluation.

In this paper, we will focus mobile learning specially on the use of mobile devices (i.e. tablets or smart phones) when developing CSCL activities. In this sense, mobile learning educational process can be considered as any learning and teaching activity that is possible through mobile tools or in settings where mobile equipment is available [18]. Therefore, we consider that mobile devices do not change significantly the CSCL processes and methodologies presented in the next sections. Hence, for the sake of simplicity, in the rest of the paper we will refer to online collaborative learning or CSCL only, which implicitly include MCSCL and collaborative learning supported by mobile devices.

2.3. Trustworthiness Models and Normalization

According to [19], there is a degree of convergence on the definition of trustworthiness. This can be summarized as follows: trustworthiness is a particular level of the subjective probability with which an agent assesses that another agent (or group of agents) will perform a particular action, before the agent can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects its own action. Regarding trustworthiness and e-Learning, according to [20], a trustworthy e-Learning system is a learning system, which contains reliable serving peers and useful learning resources.

As stated by the authors in [21], through the study of the most relevant existing trust models, trustworthiness modelling can be classified into trustworthiness assessment and prediction models (note that in the literature on trustworthiness modelling, the terms determination and estimation are also used to refer assessment and prediction respectively). The first formally trustworthiness model related to information technology services was proposed in [22] from three levels. This approach considers the main factors and rules dealing with trustworthiness, which can be summarized as follows:

1. Basic trust is the general trusting disposition of an agent at time.
2. General trust represents the trust that agent has on other agent at time.
3. Situational trust is the amount of trust taking into account a specific situation.

It is worth mentioning that this early proposal takes into account the time factor (discussed in Section 2.5) as a key trustworthiness component in the model.

Although trustworthiness models can be defined and included as a service in e-assessment security frameworks, there are multiple issues related to trustworthiness, which cannot be managed without normalization [23]. Among these issues, we can highlight trustworthiness multiple sources, different data formats, measure techniques, and other trustworthiness issues, such as rules, evolution, or context. Hence, in [13], we justify why trustworthiness normalization is needed and a normalized trustworthiness model is proposed by reviewing existing normalization procedures for trustworthy values applied to e-assessments.

2.4. Trustworthiness and Information Security

To overcome security deficiencies discussed above, we researched into enhancing technological security models with functional approaches [11, 12, 13]. In [20], a trustworthy e-Learning system is defined as a learning system, which contains reliable serving peers and useful learning resources. As stated by the authors in [21], through the study of the most relevant existing trust models, trustworthiness modelling can be classified into trustworthiness assessment and prediction models. In this paper, we considered both purposes of trustworthiness. In addition, we also consider trustworthiness models, rules, factors and features that we discussed in [11, 12, 13] with the aim to enhance security in e-Learning through trustworthiness methods.

To establish the difference between assessment and prediction, in [21] it is stated that trustworthiness prediction, unlike trust assessment, deals with uncertainty as it aims to predict the trust value over a period in the future. In such cases, the accuracy of the trust values at a point in time in the future is an important issue to be considered, as the future of business decisions will be based on these.

2.5. Time Factor and Trustworthiness Sequences

Several studies investigating trustworthiness show that time factor is strongly related to trustworthiness [20, 24, 25]. The authors in [20] stated that trust is dynamic and will attenuate when time goes by. For instance, A trusts B at time t_0, but A might not trust B in a follow-up time t_1. In [23], it is presented the design and development of a trust management system. This system addresses its specifications and architecture to facilitate the system implementation through a module-oriented architecture. Among the modules of the system, the authors define a module for dynamic assessment, which includes trust levels assessment based on dynamic trust criteria. The module integrates assessment from all parts to calculate trust value by the weighted average.

As aforementioned, we can consider both assessment and prediction trustworthiness models. Although the models reviewed analysing trustworthiness include the time factor as a key component, we need further modelling techniques that allow us to conduct trustworthiness assessment towards prediction. To this end, we reviewed the concept of Trustworthiness History Sequence [25]. In the context of grid services, Trustworthy History Sequence is a history record of trustworthy of grid service that the requester has traded with. It can be denoted with an ordered tuple where each component is the trustworthiness score of the transaction between a requester and a service.

2.6. Predicting Trustworthiness

Trustworthiness prediction models, to the best of our knowledge, have been little investigated in the context of e-assessment, even in a general prediction scope. The existing literature suggests that the term trust prediction is used synonymously and interchangeably with the trust assessment process [21] presented in the sections above. Moreover, trustworthiness does not focus on an isolated technical application, but on the social context in which it is embedded. Although trustworthiness building can be supported by institutions, there is no easy way out [26]. In addition, the building of trust can be a very lengthy process, the outcome of which is very hard to predict.

Several studies investigating trustworthiness prediction were carried out with neural networks [21, 25, 27]. In [21], the authors propose the use of neural networks to predict the trust values for any given entities. The neural networks are considered one of the most reliable methods for predicting values [21]. A neural network can capture any type of non-linear relationship between input and output data through iterative training, which produces better prediction accuracy in any domain such as time series prediction. The key contribution of this work is focused on the dynamic nature of trust, in which the performance of this approach is tested under four different types of data sets (e.g. non-uniform stationary data, different size, etc.), and the optimal configuration of the neural network is identified.

In [25], the authors stated that trustworthiness prediction with the method of neural network is feasible. The experiments presented in [25] confirm that the methods with neural networks are effective to predict trustworthiness. This method is based on defining a neural network structure, a neural network constructing, an input standardization, a training sample constructing, and the procedure of predicting trustworthiness with trained neural network.

The work presented in [27] proposes a novel application of neural network in evaluating multiple recommendations of various trust standards. This contribution presents the design of a trust model to derive recommendation trust from heterogeneous agents. The experimental results show that the model has robust performance when there is high prediction accuracy requirement or when there are deceptive recommendations.

Moreover, other trustworthiness models were proposed without neural networks methods [28, 29], such as similarity approaches. In [29], it is stated that predicting trust among the agents is of great importance to various open distributed settings. The author focuses the study on peer-to-peer systems in that dishonest agents can easily join the system and achieve their goals by circumventing agreed rules, or gaining unfair advantages. These cases are closely related to e-assessment regarding anomalous assessment processes as well as integrity and identity security properties. To this end, this work proposed a trust prediction approach to capture dynamic behaviour of the target agent by identifying features, which are capable of describing context of a transaction. A further work [28], on users' rating systems, presents experimental results which demonstrate that ratings volume is positively associated with trust, as well as the congruence between one's own and others' opinions. This study also demonstrates that rating source

and volume interact to impact credibility perceptions, reliance on user-generated information, and opinion congruence. These results indicate important theoretical extensions by demonstrating that social information may be filtered through signals indicating its veracity, which may not apply equally to all social users.

2.7. Previous Trustworthiness Methodological Approaches

To date, little research has been carried out to build trustworthiness methodological approaches. However, in the context of business processes, the authors in [30] propose a generic methodology, called Trustworthiness Measurement Methodology (TMM). This methodology can be used to determine both the quality of service of a given provider and the quality of product. The scope of this study is the business processes, but the key concept of this methodology is the interaction between agents. Indeed, this is the same topic that we study in collaborative learning, but in our context, considering students' interactions and trustworthiness between them. This methodology is based on the following phases:

1. Determine the context of interaction between the trusting agent and the trusted entity.
2. Determine the criteria involved in the interaction.
3. Develop a criterion assessment policy for each criterion involved in the interaction.
4. Determine the trustworthiness value of the trusted entity in the given context.

In [31], the authors presented the foundations of formal models for trust in global information security environments, with the aim of underpinning the use of trustworthiness based security mechanisms as an alternative to the traditional ones. As stated by the authors, this formal model is based on a novel notion of trust structures, which is built on concepts from trust management and domain theory as well as features at the same time a trust and an information partial order. The formal model is focused on the following target aspects:

1. Trustworthiness involves entities.
2. Trustworthiness has a degree.
3. Trustworthiness is based on observations.
4. Trustworthiness determines the interaction among entities.

In addition to the methodology and formal approaches, in another work [32], a trust architecture is presented by introducing a basic trust management model.

3. Trustworthiness and Security Methodology Approach

In this section, we first describe the main theoretical features of our methodological approach and then, the summary of its key phases is presented. Finally, we detail each phase by analysing the processes, data, and components involved in the methodology.

3.1. Theoretical Analysis

In these sections, we present our methodological approach called Trustworthiness and Security Methodology (TSM) in CSCL. TSM is a theoretical approach devoted to offering a guideline for designing and managing security in mobile collaborative e-Learning activities through trustworthiness assessment and prediction.

TSM is defined in terms of TSM cycles and phases, as well as, components, trustworthiness data and main processes involved in data management and design. We define a TSM phase as a set of processes, components, and data. TSM phases are sequentially arranged and the three main phases (see Fig. 1) in TSM form a TSM design and deploy cycle (i.e. TSM-cycle). Each TSM-cycle corresponds to an interaction over the overall design process. Firstly, these concepts are presented as a methodological approach and then we complete the theoretical analysis with those methods and evaluation processes that we discussed in our previous research [11, 12, 13].

TSM aims to deliver solutions for e-Learning designers. TSM supports all analysis, design, and management activities in the context of trustworthiness collaborative learning activities, reaching security levels defined as a part of the methodology. Therefore, TSM tackles the problem of security in CSCL through the following guidelines and main goals:

1. Define security properties and services required by e-Learning designers.
2. Build secure CSCL activities and to design them in terms of trustworthiness.
3. Manage trustworthiness in learning systems with the aim of modelling, predicting, and processing trustworthiness levels.
4. Detect security events which can be defined as a condition that can violate a security property, thus introducing a security breach in the learning system.

The scope of our methodological approach is an e-Learning system formed by collaborative activities developed in a Learning Management System (LMS). The LMS has to provide support to carry out these activities and to collect trustworthiness data generated by learning and collaboration processes. Although in the context of collaborative e-Learning we can consider several actors with different roles in the overall process, for the sake of simplicity, we only consider the most significant actors and roles related to this research, as follows:

1. Students, as the main actors in the collaborative learning process and as targets of the trustworthiness analysis.
2. Designers, that represent the role in charge of all e-Learning analysis and design tasks.
3. Managers, that develop management processes, such as deployment, monitoring or control tasks.

3.2. Methodology Key Phases

As shown in Fig. 1, the TSM methodology is divided into three sequential phases:

1. Building Trustworthiness Components, integrated into the design of secure collaborative learning activities.
2. Trustworthiness Analysis and Data Processing, based on trustworthiness modelling.
3. Trustworthiness Assessment and Prediction, to detect security events and to refine the design process.

Although we assess each phase of the methodology as potential sets of concurrent processes (see next sections), these core phases have to be developed following the sequential phases presented. The main reason for defining this sequential model is the input and output flow. In other words, the output of one phase is the input of the next one. For instance, we can only start the data collection phase when trustworthiness components are deployed. Likewise, we cannot start trustworthiness prediction or assessment until data processing has been completed.

Despite the sequential model between each phase, we can consider the overall process, formed by these three phase, as a TSM-cycle. Each TSM-cycle allows e-Learning designers to improve the collaborative learning activities from the results, and trustworthiness decision information retrieved from the previous cycle. This information can introduce design enhances which will be deployed in the next deployment (i.e. the next time that the students will carry out the activity supported by the learning component). In terms of the data flow between TSM-cycles, the input for the new design iteration is the trustworthiness decision information. For instance, if decision information shows that there exists a deficiency in a component, the detected impediment can be overcome through design changes that are deployed in the next TSM-cycle execution.

3.3. Building Trustworthiness Components

The first phase of TSM deals with the design of collaborative activities. The key challenge of the design process is to integrate trustworthiness data collection inside the learning process. In other words, the trustworthiness component has to carry out its learning purpose. In addition, the learning component has to produce trustworthiness basic data. Moreover, data collection methods and processes should not disturb the learning activity. To this end, we propose the processes, data, and components that can be seen from the diagram in Fig. 2. Due to the first goal of the methodology is to design the trustworthiness component, we divide this phase into the following analysis considerations:

1. Collaborative learning activities generate a significant amount of interactions. Due to students' interactions are closely related to trustworthiness modelling, designers have to consider and analyse each interaction, which may be related to trustworthiness.
2. Analyse and determine relations between students' interaction and trustworthiness could be a challenging task in e-Learning design. Hence, we propose the study of trustworthiness factors [11], which can be defined as those behaviours that reduce or build trustworthiness in a collaborative group. Trustworthiness factors can be divided into trustworthiness reducing factors and trustworthiness building factors. This resource will allow designers to determine those interactions, which may generate trustworthiness basic data.

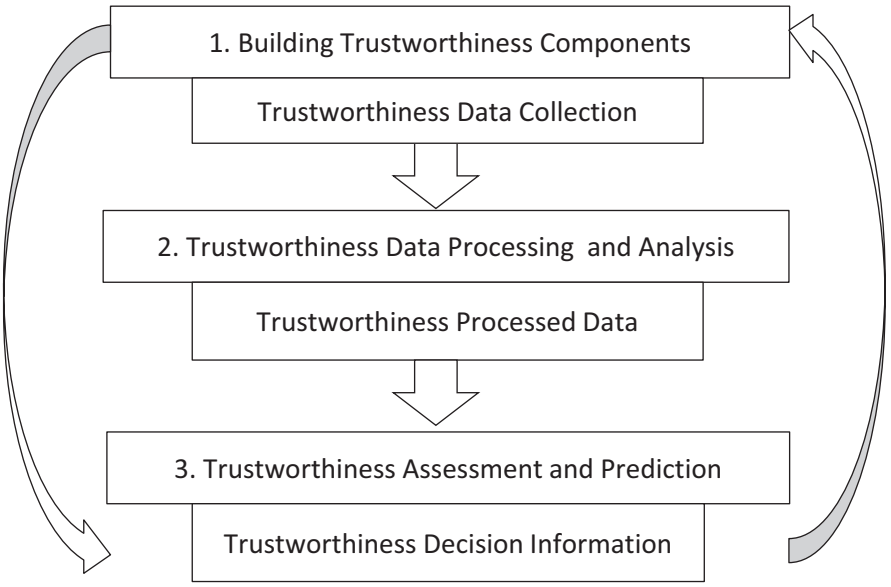

Figure 1: TSM Key Phases

3. Designers have to model security issues so that they are compatible with trustworthiness data and students' interactions.

Based on the above considerations, we propose the analysis of general security properties and services presented in [4]. Through selecting and analysing security properties we can connect trustworthiness, interactions, and security requirements in terms of collaborative learning activities.

From the study of security properties, students' interactions and trustworthiness factors, the initial collaborative learning activity has evolved to a peer-to-peer assessment component. Once we endowed the collaborative activity with security and trustworthiness, the next process is focused on data collection. To this end, we define research instruments for data collection intended to retrieve all trustworthiness data generated by the peer-to-peer assessment component.

Note that, for the sake of simplicity, we present a case dealing with one collaborative activity only, which generate its peer-to-peer assessment component. Despite this, the case may be extended to a set of collaborative activities implemented in one or several peer-to-peer components. Moreover, the components can be supported by several research instruments or a peer-to-peer component, including multiple collaborative activities. Eventually, the result in any case (i.e. single and multiple activities, components and instruments) is a set of trustworthiness basic data that will feed the next phase of the methodology. For this reason, we define the input of the next phase in terms of multiple trustworthiness data sources.

We suggested the need of modelling activities, components, security properties, or interactions in the context of a general design process. This process may be a challenge if the e-Learning designer does not use suitable modelling tools. To overcome this impediment, we reviewed the Educational Modelling Language (EML) [33] that, with the indications presented in [4], allows designers to tackle with modelling security, CSCL activities and interactions.

3.4. Trustworthiness Analysis and Data Processing

So far, the e-Learning designer has built the trustworthiness component, which will be deployed in the LMS. It is worth mentioning that the deployment of collaborative learning activities may involve multiple LMSs. In fact, we are proposing a learning activity deployment in conjunction with research instruments for data collection. The implementation of these instruments may require additional technological solutions such as normalization processes. Trustworthiness modelling and normalization processes in TSM (see Fig. 3) are based on the key concepts presented

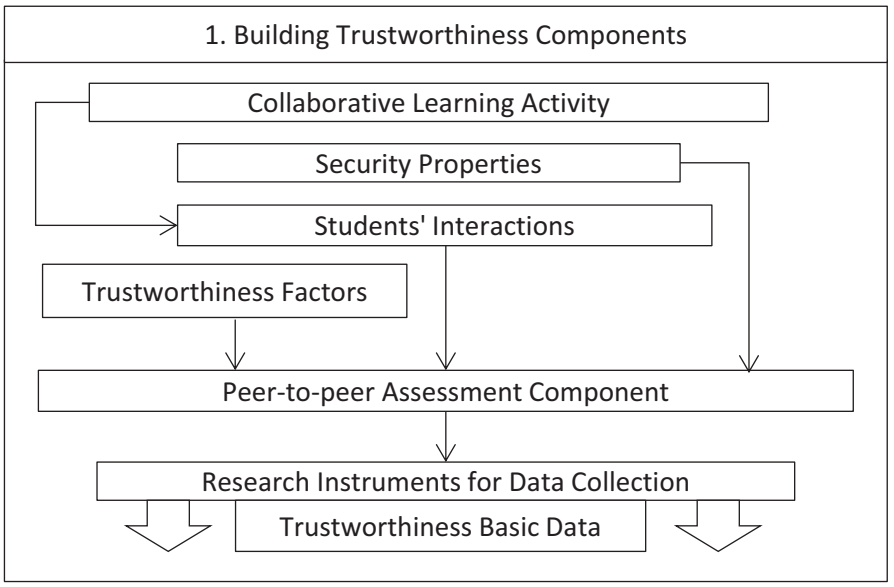

Figure 2: Phase 1: Building Trustworthiness Components

in the rest of this subsection (further information and details of these concepts can be found in our previous research [11, 12, 13]).

We introduced the concept of Trustworthiness Indicator as a measure of trustworthiness factors. Trustworthiness factors were presented (see Section 3.3) as those behaviours that reduce or build trustworthiness in a collaborative activity and they were integrated in the design of research instruments. Therefore, we define a Trustworthiness Indicator as a basic measure of a trustworthiness factor that is implemented by a research instrument and integrated in the peer-to-peer assessment component. Finally, Trustworthiness Levels can be defined as a composition of trustworthiness indicators. The concept of levels is needed because trustworthiness rules and characteristics must be considered and, consequently, we have to compose this more complex measure [11].

Regarding normalization functions there are several reasons that impede the management and processing of trustworthiness levels directly. Among them, we can highlight several aspects, such as multiple sources, different data formats, measure techniques and other trustworthiness factors such as rules, trustworthiness evolution, or context. Therefore, both trustworthiness indicators and levels have to be normalized through normalization functions. The selection of these functions depends on the data sources and the format selected for each instrument for data collection [13].

Once trustworthiness modelling concepts are defined, the task of data processing starts, and then basic data from trustworthiness data sources is computed in order to determine indicators or levels, for each student, group of students, evaluation components, etc. The main challenge of data processing in this case is that extracting and structuring these data are a prerequisite for trustworthiness data processing. In addition, with regarding to computational complexity, extracting and structuring trustworthiness data is a costly process. Moreover, the amount of basic data tends to be very large [12]. Therefore, techniques to speed and scale up the structuring and processing of trustworthiness basic data are required (see [12] for a parallel implementation approach to be developed in the context of trustworthiness data processing).

3.5. Trustworthiness Assessment and Prediction

From the trustworthiness data computed in the previous phase, we can carry out both assessment and prediction processes, which allow e-Learning managers to make security decisions based on the output of this phase (i.e. trustworthiness decision information). Furthermore, this information can be taken into account as input data for an iterative design process as mentioned in Section 3.2.

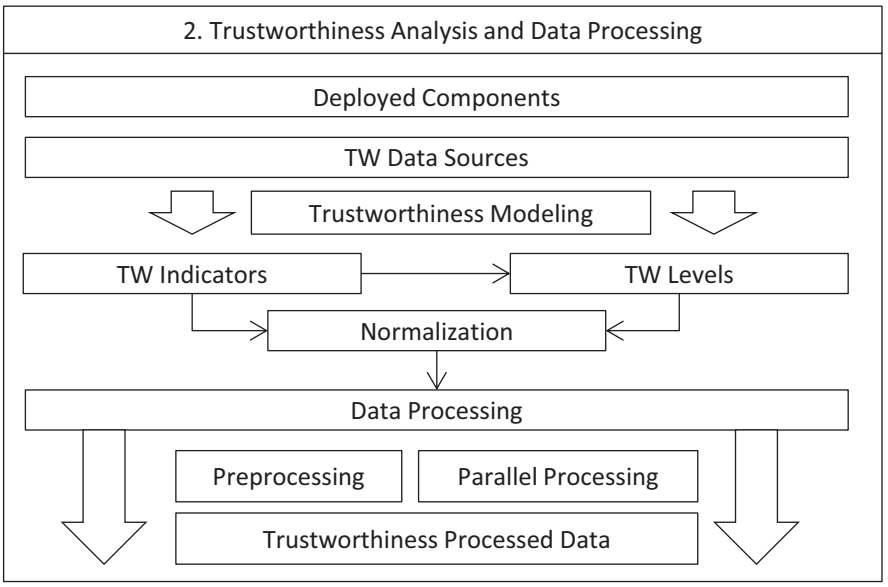

Figure 3: Phase 2: Trustworthiness Analysis and Data Processing

Trustworthiness assessment and prediction stem from the analysis of the time factor in trustworthiness. Fig. 4 shows how trustworthiness assessment and prediction begin with the conversion of processed data into trustworthiness sequences by considering the time factor. The concept of trustworthiness sequence is related to levels and indicators and can be defined as the ordered list of a student's trustworthiness normalized levels when the student is performing the peer-to-peer assessment component over several points in time.

Once trustworthiness sequences are built, the e-Learning manager is able to set out predictions and assessment processes. As presented in [11], methods intended to predict and assess trustworthiness are available in the context of peer-to-peer assessment. The e-Learning designer has to select and determine suitable methods for the specific target scenario.

We cannot use trustworthiness decision information (i.e. reliable trustworthiness information) without the validation process. The validation process is intended to filter anomalous cases, to compare results that represent the same information from different sources, and to verify results using methods such a similarity coefficients. Nevertheless, this information may indicate signs and the complex nature of trustworthiness modelling requires additional validation processes. These validation models can be classified into internal and external, and each type may involve automatic and manual tasks. For instance, in the context of e-assessment, we could compare trustworthiness results generated by the peer-to-peer assessment component to external (respect to the peer-to-peer component) results from the manual tutor evaluation. Moreover, this comparison could be automatically developed by the system and analysed by the tutor before taking any decision.

Finally, trustworthiness decision information is available and then e-Learning managers can analyse valid and useful information devoted to reporting security events, improve the framework design, or manage security enhances. In the rest of the paper we present specific TSM aspects in real online courses, focused on trustworthiness assessment (see Section 4) and trustworthiness prediction (see Section 5).

4. Trustworthiness Methodology Evaluation

In order to evaluate and support the application and deployment of TSM, in this section, we concrete several significant aspects of TSM. These aspects are considered in terms of specific methods and techniques through their application in real online courses.

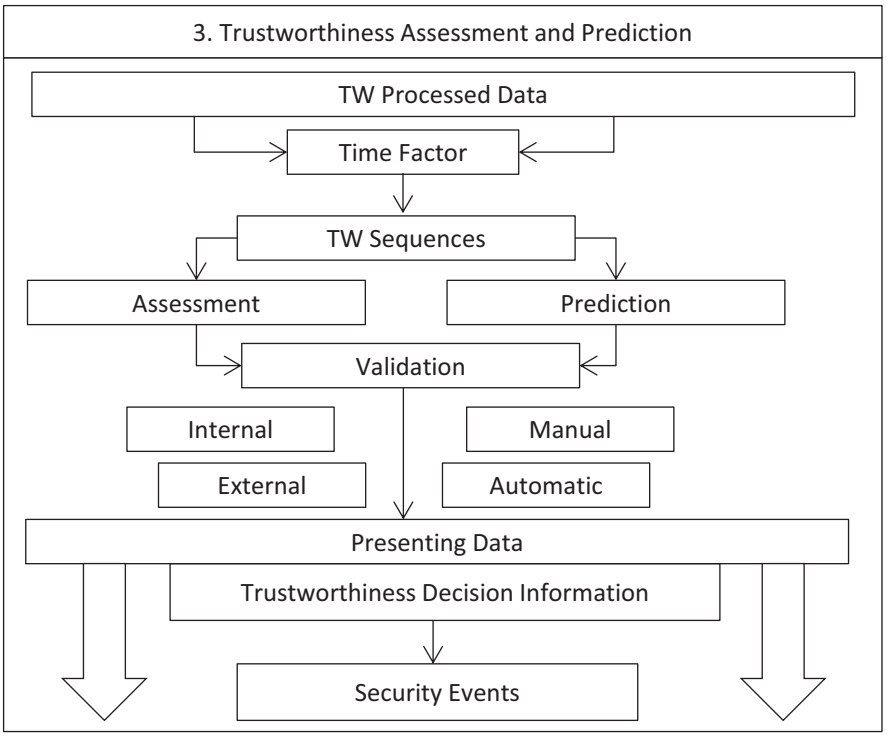

Figure 4: Phase 3: Trustworthiness Assessment and Prediction

4.1. Real Online Courses

We carried out two studies [12, 34] based on real online courses at the Open University of Catalonia[1]. These studies were performed with the aim to experiment with specific trustworthiness methods and techniques involved in TSM as well as to illustrate specific applications and to evaluate the feasibility of the TSM.

In the first study [12], the mobile collaborative activities represented a relevant component of the e-assessment of the course. Students' evaluation was based on a hybrid continuous evaluation model by using several manual and automatic evaluation instruments. There were 12 students distributed in three groups and the course was arranged in four stages. These stages were taken as time references in order to implement trustworthiness sequences. At the end of each collaborative stage, each student had to complete a survey. The coordinator of the group had to complete two reports, public and private, and at the end of each stage, the members the group was evaluated by the coordinator. General e-Learning activities were supported by a standard LMS, which offered both ratings systems and general learning management indicators. Given the low number of students, we could study the data in much more detail and flexibility. Likewise, we could experiment with several design alternatives and adapting the model to the design cycles proposed in TSM (see Section 3).

The second study [34] extended the scope of the first one to a more standard scenario in which we could not manage so much flexibility and manual processes. The course was focused on peer-to-peer e-assessment and it has the following main features:

- Students' assessment was based on a continuous assessment model by using several manual assessment instruments. Manual assessment was completed with automatic methods, which represented up to 20% of the total student's grade. Therefore, we implemented a hybrid assessment method, which combined manual and automatics assessment methods, and the model allowed us to compare results in both models.

[1] http://www.uoc.edu

- Number of students participating: 12 students performed a subjective peer-to-peer assessment, that is, each student could assess any student in the classroom following the assessment design.
- The course followed seven stages that could be taken as time references in order to validate and to analyse results. Each stage corresponded to a module of the course, which had a learning module (i.e. book) that the student must study before developing the assessment activities of the course.

From the above base course features, we built the peer-to-peer e-assessment activity encapsulated as a Continuous Assessment (CA), which was formed by three assessment activities (described in the rest of this section). Once the student has studied a module, the student receives an invitation to answer (i.e. a short text response) a set of evaluation questions about the current module. This is the first activity of the CA named the Module Questionnaire and denoted by Q. The student did not have to answer as soon as Q was sent, because the second activity of the CA was a students' forum (F) intended to create a mobile collaborative framework devoted to enhancing responses in activity Q, in other words, Q and F activities are concurrent tasks. The final activity was the core of the peer-to-peer assessment and the student has to complete a survey (P) which contained the set of responses from Q. The student had to assess each classmate's responses in Q and, furthermore, the activity of each student in the forum F was assessed. These collaborative activities were designed considering the use of mobile devices.

4.2. Building Collaborative Components with TSM

After the experience designing components in the first study, in the second one, we built a comprehensive peer-to-peer assessment component. We selected integrity and identity as target security properties for the component and, after the analysis of potential students' interactions in basic activities, the first version of the peer-to-peer assessment component was proposed.

The final version of the component had three stages: Once the student had studied a module, the student received an invitation to a survey (S1) with questions about the current module. Students did not have to answer S1 as soon as the invitation was received. The second activity of the component was a students' forum (F), which created a collaborative framework devoted to enhancing responses' quality in S1. Eventually, the student had to complete another survey (S2), which contained the set of responses over the first one (S1). By using S2, the student had to evaluate each classmate's responses as well as the participation of each student in the forum F. The design of this activity endorsed our proposal regarding the analysis of security properties, students' interactions, and factors.

Regarding research instruments and data collection, we included the following instruments:

1. Surveys.
2. Ratings.
3. Students reports.
4. LMS indicators.

To sum up, each instrument was integrated into the mobile collaborative activity (through mobile tools) and it managed its own data formats.

4.3. Notation and Terminology in TSM

Before the analysis and data processing phase, we introduce the key terms presented in the next sections (see Table 1).

4.4. Analysis and Data Processing with TSM

We analysed research instruments data formats in terms of data sources in TSM. For each case, we selected a set of normalization functions intended to convert basic trustworthiness data in normalized trustworthiness values. Normalization functions are combined with trustworthiness levels and indicators. As an example of this combination, when a student evaluates every classmate's responses, we use the following normalization function [13]:

$$N\left(tw_{R_{q,m,s}}\right) = \sum_{j=1}^{N_S} \frac{tw_{R_{q,m,j}}}{N_S - 1}, j \neq s, N_S = |S|, q \in Q, m \in M, s \in S, j \in S \quad (1)$$

Table 1: Notation and Terminology

tw_i	A trustworthiness indicator tw_i as a measure of trustworthiness factors.
$i \in I$	The set of trustworthiness indicators.
N_I	The number of trustworthiness indicators.
$m \in M$	A module m in the set of modules M.
N_M	The number of modules.
$q \in Q$	A question q in the set of questions Q.
N_Q	The number of questions.
$s \in S$	A student s in the set of students S.
N_S	The number of students.
DS_{ca}	The Continuous Assessment (CA) Data Sources, $ca \in \{R, F, Q_r, Q_c\}$.
DS_{Q_r}	The questionnaire DS for the students' responses.
DS_{Q_c}	The questionnaire DS for the number of responses.
DS_R	The peer-to-peer questionnaire DS for the score that a student has assessed a student's response.
DS_F	The forum participation DS for the number of posts.
$N()$	Normalization function to convert basic indicators in normalized trustworthiness values.
w_i	The component normalization weight for the indicator tw_i, $w_i \in (w_1, \ldots, w_n)$.
$N_2()$	Normalization function for responses data source DS_R.
$N_4()$	Normalization function for forum participation data source DS_F.
$tw_{ca_{q,m,s}}$	Trustworthiness indicator for the Continuous Assessment (CA) component.
$tw^N_{ca_{q,m,s}}$	Normalized trustworthiness indicator for the CA component.
$tw_{R_{q,m,s}}$	The trustworthiness indicator for the students' responses score data source DS_R.
$tw_{F,m,s}$	The trustworthiness indicator for the forum participation.
L^N_I	The generic normalized trustworthiness level.
$L^N_{R,m,s}$	The normalized trustworthiness level for students' responses.
$L^N_{F,m,s}$	The normalized trustworthiness level for forum participation.
$L^N_{m,s}$	The overall normalized trustworthiness level.
$CATS_s$	The Continuous Assessment Trustworthiness Sequence (CATS) ordered list.
$CATS$	The CATS matrix.
$CATS^a_s$	The active CA trustworthiness history sequence.
$CATS^c_s$	The constrictive trustworthy history.
$CATS^W_s$	The trustworthiness window sequence.

where $tw_{R_{q,m,s}}$ is the responses (R) indicator, s is the target student (i.e. the student evaluated), N_S is the number of students in the course, and q is the one of the questions evaluated in the module m.

With respect to trustworthiness normalized levels Ltw^N, we managed several indicators composition. The most suitable level in both courses is based on a weight model:

$$Ltw^N = \sum_{i=1}^{N_I} \frac{tw_i \cdot w_i}{N_I}, i \in I, w_i \in (w_1, \ldots, w_{N_I}), \sum_{i=1}^{N_I} w_i = 1, N_I = |I| \quad (2)$$

where N_I is the total number of trustworthiness indicators and w_i is the weight for the normalized indicator tw_i.

Regarding data processing, we experimented with sequential and parallel implementations [12]. Sequential approaches were feasible to manage data sources from several activities, such as responses in a survey or number of posts in a forum. However, processing the log data took too long to complete and it had to be done offline (i.e. after the completion of the learning activity). For this reason, we endowed our trustworthiness framework with parallel processing facilities.

To this end, we designed a MapReduce algorithm [12] implemented in an Apache Hadoop[2] and deployed in the

[2]http://hadoop.apache.org

RDlab[3] computing cluster. Using this model, a considerable speed up was achieved in processing large log file, namely, more than 75% for 10 nodes (see [12] for the whole results).

4.5. Assessment, Prediction and Evaluation with TSM

Peer-to-peer components were designed considering the time factor. Activities are arranged in stages that conduct the definition of trustworthiness sequences. In both studies, trustworthiness indicators and levels are instanced in points of time (e.g. the same indicator measured for each module) and arranged in trustworthiness sequences. The concept of trustworthiness sequence in an e-assessment component allows us to support assessment and prediction. Actually, it could be directly incorporated, in some cases, as input for assessment and prediction methods. Regarding validation, we experimented with a hybrid validation approach by combining manual, automatic, external, and internal validation methods. As an example of this model, we analysed similarity between manual evaluation results and automatic trustworthiness levels. The method to tackle similarity proposed is based on Pearson correlation [35].

Finally, we consider two different methods to deal with prediction. The first approach is based on neural networks [21] and the second one on collaborative filtering. On the one hand, a neural network captures any type of non-linear relationship between input and output. In our case, the input is the trustworthiness history sequence and the output is the prediction calculated by the neural network (i.e. trustworthiness predicted value). On the other hand, filtering recommendation algorithms concern the prediction of the target user's assessment, for the target item that the user has not given the rating, based on the users' ratings on observed items. In our context, items involved in the recommendation system are the students themselves.

In the rest of this paper, we focus the validation of TSM on trustworthiness prediction based on a neural network approach. Furthermore, the methods presented in this section (i.e. trustworthiness data sources, indicators, normalizations processes, and history sequences) are also applied from the view of trustworthiness prediction.

5. Evaluation of Trustworthiness Prediction

In this section, a trustworthiness prediction model is presented in the context of the real online course based on peer-to-peer e-assessment described in Section 4.1.

5.1. Normalizing Trustworthiness Data Sources

Once the peer-to-peer e-assessment has been designed, we analyse and define trustworthiness data sources and levels. In the context of Continuous Assessment (CA), we defined a trustworthiness data source as those data generated by the CA that we use to define trustworthiness levels as presented in [11, 12, 13]. Each CA correspond to a module $m \in M$, which is a unit of the course. The modules will be used as a point in time references. Each CA (i.e. one CA per module) will manage three data sources, which are denoted with the following ordered tuples:

$$DS_{Q_C} = (M, Q, S, count) \tag{3}$$

where the questionnaire data source DS_{Q_C} is defined as the total number of responses (*count*) that each student in S has answered in the questionnaire Q for the module M.

$$DS_{Q_R} = (M, Q, S, res) \tag{4}$$

where the questionnaire data source DS_{Q_R} is defined as the response *res* (i.e. a student answers *res* to a question) that each student in S has responded regarding a specific question in Q in the module M.

$$DS_F = (M, F, S, count) \tag{5}$$

where the forum participation data source DS_F is defined as the total number of posts (*count*) that each student in S sent to a forum F regarding a specific question in Q in the module M.

[3] http://rdlab.lsi.upc.edu

$$DS_R = (M, Q, S, SS, score) \tag{6}$$

where the responses data source denotes the score that a student (in S) has assessed a student's (in SS) response of a question in Q. Hence, S is the set of students who assess and SS is the set of students who are assessed by students in S.

In this case, modelling trustworthiness involves multiple complex and heterogeneous data sources with different formatting, which cannot be managed without normalization. According to the model presented in [13], we define a normalized trustworthiness indicator for the case of a CA as follow:

$$tw^N_{ca_{q,m,s}} = N\left(tw_{ca_{q,m,s}}\right), ca \in DS_{R,F,Q_r,Q_c}, q \in Q, m \in M, s \in S \tag{7}$$

where DS_{R,F,Q_r,Q_c} are the CA data sources, S is the set of students, M is the set of modules, and Q is the set of questions in each module.

We now define the normalization functions. Note that although in [13] we included four normalization functions, in this case, a subset is selected: N_2 and N_4. The reason for this is that we focus the data analysis on two data sources, forum participation (N_4) and questionnaires (N_2). Regarding the responses data source R, a student can assess every classmate's responses. To this end, we use the normalization function N_2:

$$N_2\left(tw_{R_{q,m,s}}\right) = \sum_{i=1}^{N_S} \frac{tw_{R_{q,m,i}}}{N_S - 1}, i \neq s \tag{8}$$

where $tw_{R_{q,m,s}}$ is the responses indicator, s is the target student (i.e. the student who is assessed), N_S is the number of students in the course, and q is the one of the questions assessed in the module m.

It is worth mentioning that the scale for $tw_{R_{q,m,s}}$ must be converted to integer values before normalizing with function N_2. Similarity, the forum participation indicator also needs normalization. In this case, we apply the normalization function N_4:

$$N_4\left(tw_{F,m,s}\right) = \frac{tw_{F,m,s}}{T_F}, m \in M, s \in S \tag{9}$$

where T_F is the maximum number of post in the forum by a student s in the module m.

5.2. Trustworthiness Levels and Sequences in e-Assessment

We normalize the trustworthiness indicators for forum participation and responses (i.e. a student answers a question in the questionnaire). Then, trustworthiness levels [11] are defined in order to measure students' overall trustworthiness. To this end, we define the following trustworthiness levels:

$$L^N_I = \sum_{i=1}^{N_I} \frac{\left(tw^N_i * w_i\right)}{N_I}, i \in I, w_i \in (w_1, \ldots, w_{N_I}), \sum_{i=1}^{N_I} w_i = 1 \tag{10}$$

where N_I is the total number of trustworthiness indicators and w_i is the weight assigned to tw_i.

Following this model, we first combine the trustworthiness indicators of each question in the module and then, the overall trustworthiness level for the student in a specific module $m \in M$ is defined:

$$L^N_{R,m,s} = \sum_{q=1}^{N_Q} \frac{\left(tw^N_q * w_q\right)}{N_Q}, q \in Q, N_Q = |Q|, \sum_{q=1}^{N_Q} w_q = 1, w_q = \frac{1}{N_Q}, m \in M, s \in S \tag{11}$$

$$L^N_{F,m,s} = N_4\left(tw_{F,m,s}\right), m \in M, s \in S \tag{12}$$

$$L^N_{m,s} = \sum_{j=1}^{2} \frac{\left(Ltw^N_j * w_j\right)}{2}, j \in \{L^N_{F,m}, L^N_{R,m}\}, \sum_{j=1}^{2} w_j = 1, w = (0.4, 0.6), m \in M, s \in S \tag{13}$$

where $L^N_{m,s}$ is the overall trustworthiness level for the student s in the module m, calculated by combining the trustworthiness level for responses $L^N_{R,m,s}$ and the trustworthiness level for forum participation $L^N_{F,m,s}$.

Once trustworthiness levels are defined, we endow our model with time factor. Although the concept of trustworthiness sequence was defined in the context of grid services and requesters [25], it is feasible to apply this approach to another modelling scenario such as peer-to-peer e-assessment. The only requirement is time factor, in other words, the model should allow us to compute an overall trustworthiness level referred to multiple points of time. Therefore, we define Continuous Assessment Trustworthiness Sequence CATS as the ordered list of a student's trustworthiness history levels over several points in time:

$$CATS_s = \left(L^N_{m_1,s}, \ldots, L^N_{m_k,s}, \ldots, L^N_{m_{N_M},s}\right) m_k \in M, s \in S \tag{14}$$

where M is the set of modules, each module m_k refers to a point in time and $L^N_{m_k,s}$ is the overall trustworthiness level for the student s in the module m_k.

Likewise, we can define the overall students' CA trustworthiness history sequence as the matrix:

$$CATS = \begin{pmatrix} L^N_{m_1,s_1} & \cdots & L^N_{m_1,s_{N_S}} \\ \vdots & \ddots & \vdots \\ L^N_{m_{N_M},s_1} & \cdots & L^N_{m_{N_M},s_{N_S}} \end{pmatrix} \tag{15}$$

where N_M is the number of modules (i.e. points in time analysed), and N_S is the number of students in the course.

5.3. Trustworthiness Sequences Results

Processing trustworthiness sequences results involves large amount of data generated by the peer-to-peer activity of the CA. To this end, we compute the following elements:

1. The trustworthiness history sequence matrix has $N_S * N_M, N_S = |S|, N_M = |M|$ elements.
2. For each element in $CATS$, $L^N_{m,s}$, we compute both forum participation and responses trustworthiness levels.
3. Although forum participation is a single indicator, with respect to responses, there are three different questions.
4. Moreover, for each trustworthiness levels we compute each student's score for the indicator.

With the aim of managing this trustworthiness sequences results, we developed a data parse *Java* tool called *parse_tw_tuples* that converts peer-to-peer values into basic tuples presented above. This tool generates basic tuples from the web applications and these primitive records can be imported in a relational database for further processing. In order to deal with the results, we have to consider the size of the result set of records generated by each data source. At the end of the process the responses data source maximum size is:

$$|DS_R| = |M| \times (|Q| + 1) \times |S| \times |S| \tag{16}$$

where $|M|$ is the number of modules, $|Q|$ is the number of questions (+1 is added because the student also assesses the forum activity), and $|S|$ is the number of students who could participate in both questionnaires (i.e. Q and P).

The total number of computed tuples is:

$$|DS_R| = 10.522 \tag{17}$$

To sum up, the diagram depicted in Fig. 5 shows the overall process including how we have to normalize data sources. Then, this figure shows the creation of trustworthiness indicators and levels, and finally, the procedure presented to compose trustworthiness sequences.

5.4. Predicting with Trustworthiness Sequences

So far, we have presented the design of trustworthiness history sequences in the peer-to-peer assessment components of the target online course. To this end, we have to consider the main concepts presented in [25] related to trustworthiness history sequence as a foremost step in trustworthiness prediction based on neural network design.

Active trustworthiness history sequence is the recent trustworthy history sequence. Then, we define active CA trustworthiness history sequence $CATS^a_s$ as the ordered list of students' trustworthiness levels over the points in time:

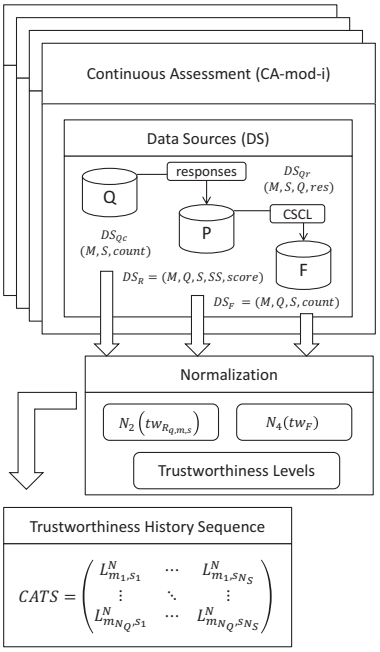

Figure 5: CA data sources, normalization and trustworthiness sequences

$$CATS_s = \left(L^N_{m_1,s}, \ldots, L^N_{m_k,s}, \ldots, L^N_{m_{N_M},s}\right), m_k \in M, s \in S \tag{18}$$

$$CATS^a_s = \left(L^N_{m_{N_Q-a+1},s}, L^N_{m_{N_Q-a+2},s}, \ldots, L^N_{m_{N_M},s}\right), s \in S \tag{19}$$

where M is the set of modules, each module m_k refers to a point in time, and $L^N_{m_k,s}$ is the overall trustworthiness level for the student s in each module.

Constrictive trustworthy history is the subsection average of active trustworthy history sequence.

$$CATS^c_s = \left(L^N_{m_{1\ldots N_S},s}, L^N_{m_{r+1\ldots N_Q},s}, \ldots\right), s \in S \tag{20}$$

where each element in the tuple is the average of a subset of elements in $CATS^a_s$, and k is the number of inputs of NN.

These tuples are presented in order to prepare those input sets that are required in neural network training and validation. The concept of trustworthiness sequences in prediction with neural networks is also suggested in [21]. In this proposal, the trustworthiness sequence is split into subsequences of fixed sizes, without average transformation:

$$CATS^W_s = \left(L^N_{m_1,s}, \ldots, L^N_{m_w,s}\right), \left(L^N_{m_{w+1},s}, \ldots, L^N_{m_{2w},s}\right), \ldots, s \in S \tag{21}$$

where each component in the trustworthiness window is a subset of the $CATS_s$.

5.5. Designing a Neural Network e-Assessment Proposal

We reviewed complementary related trustworthiness prediction work. Among existing models, we select the neural network-based approaches for predicting trust values presented in [21] and [25], because these approaches are feasible in the context of e-assessment. These models present several significant differences, especially with respect to how to build training sets, these differences are considered in our e-assessment proposal. Although we evaluated both approaches, in the rest of the paper, we address our NN design to a training model based on $CATS^W_s$. We consider this approach more suitable for our case because $CATS^W_s$ generates a greater amount of training sequences.

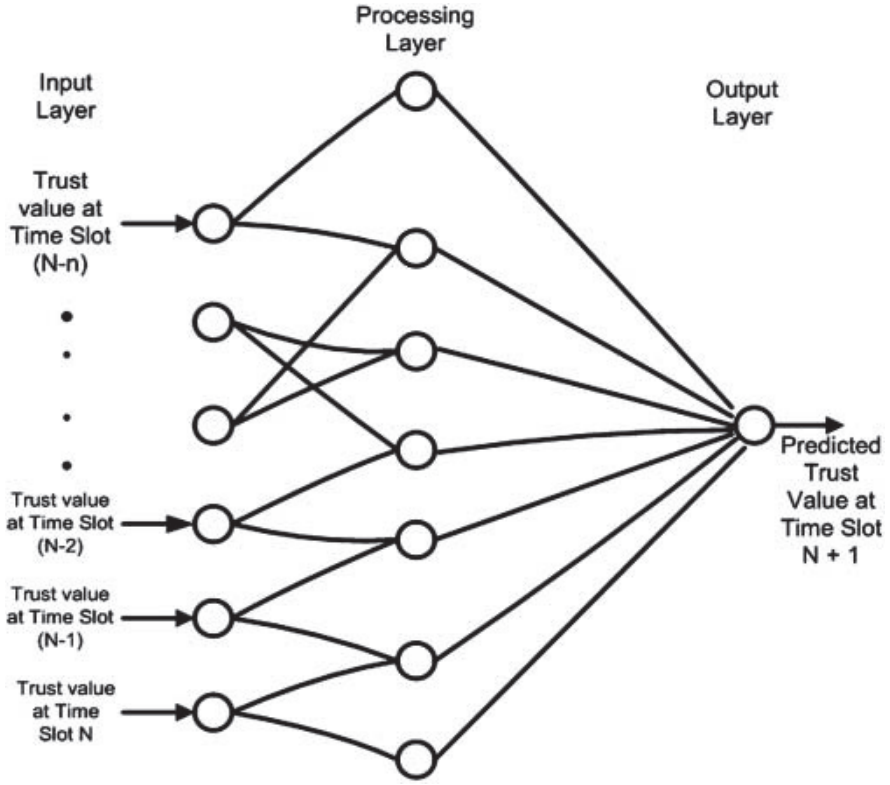

Figure 6: A simple NN approach for trust prediction [21]

A neural network can capture any type of non-linear relationship between input and output data through iterative training. In our case, the input is the CA trustworthiness history sequence formed by trustworthiness results generated by the peer-to-peer assessment component, and the output is the prediction calculated by the neural network (i.e. trustworthiness predicted value):

$$L^{N}_{m_{t+1},s} = NN\left(CATS_{s}\right), s \in S \tag{22}$$

where entity s denotes the student whose normalized trustworthiness level value is being predicted through the $CATS_{s}$ representing data generated by the peer-to-peer activity of the CA, and m_{t+1} denotes the trustworthiness point in time in the future predicted by the function NN for the student s (i.e. the output of the NN).

As presented in [21], the main principle of neural computing is the decomposition of the input-output relationship into a series of linearly separable steps using hidden layers. The NN architecture (see Fig. 6) is composed of sets of neurons that are arranged in multiple layers. The first layer, which inputs are fed to the network, is called the input layer. The last layer, which produces the NN output, is called the output layer. The layers in between these two layers (i.e. between input and output layers) are all hidden layers. The input consists of values that constitute the inputs for the hidden layers.

Every node computes a weighted function of its inputs and applies an activation function to compute the next output. The output is transmitted to all the connected nodes on the next layer with associated weights. The activation of each node depends on the bias of the node, which calculates the output as follows:

$$y_{j} = \sum_{i=0}^{n} w_{ij} x_{i} \tag{23}$$

where y is the result of the summation of the product of the input x with its associated interconnection weight w. The initial weights are assigned randomly but are gradually changed to reduce the error. The difference between the

desired output and the actual output constitutes the input to the back propagation algorithm for training the network based on the difference.

Through the iterative training, the NN produces better prediction accuracy in the domain of time series prediction, such as trustworthiness history sequences.

5.6. Simulation and Analysis of Results

With the aim of implementing the NN for trustworthiness prediction, we evaluated several simulators. Among them, we selected *Emergent*[4] as a suitable software tool that reaches all the requirements for our case. Emergent (formerly PDP++) is defined as a comprehensive, full-featured deep neural network simulator that enables the creation and analysis of complex, sophisticated models [36]. The main reasons to use *Emergent* in the context of this paper can be summarized as follow:

- *Emergent* provides powerful visualization and infrastructure tools.
- Provides a structured environment for using and modify models based on NN templates, as well as, test and training programmes.
- *Emergent* is completely open source software.
- Highly optimized runtime performance. In fact, we deployed the simulator environment in a virtual machine running on a personal computer.

With the aim of developing a first simulation approach in *Emergent*, we carried out the following tasks:

1. A new simulation project was created based on the template BpStd (i.e. standard initialization of back-propagation). This resource is provided by *Emergent* and allows the designer to begin the neural network design from a standard configuration.
2. As shown in Fig. 7 we generated and configured a standard network, specifying number of layers, layer names, sizes, types, and connectivity. The NN is formed by 3 layers with 2 input values and 1 output. In terms of *Emergent* design, the geometry for both input and hidden layers is a 2 units x 1 units matrix.
3. The NN geometry corresponds to the size of the data contained in the *StdInputData* table. This table contains each student's trustworthiness window sequence $CATS_s^W$ defined in Section 5.4. The data import process was managed through text file elements (see Fig. 8). *Emergent* offers import and export tools that bind the *StdInputData* tables and the text files.
4. Once NN basic design and input data were configured, the next step was the training process of the NN. Following the model defined in Section 5.4, we split the input values for each student into two trustworthiness sequences (i.e. training and test). The training trustworthiness window sequence contained 5 instances (i.e. time slots or modules in the course), which were arranged in tuples of 3 elements. The 3-tuple was also divided into the input values and the output result. Therefore, for each $CATS_s^W$ sequence we generate tuples of 3 elements containing the 2 input values and the output expected value.
5. The training process is managed by *Emergent* in the *BpTrain* programme whose initial parameters are shown in Fig. 9.
6. Finally, we introduced the test elements in order to validate the model.

The deviation in prediction results for each student is depicted in Fig. 10. The sample of the experiment was formed by 12 students. Fig. 10 presents the results obtained from the NN simulation process for each student. The horizontal axis represents students and the vertical axis represents the difference between the value predicted by the NN and the test value (i.e. the prediction error in absolute value). For instance, the NN for the student 5 predicted a value with a 2.54% of error.

[4]https://grey.colorado.edu/emergent/

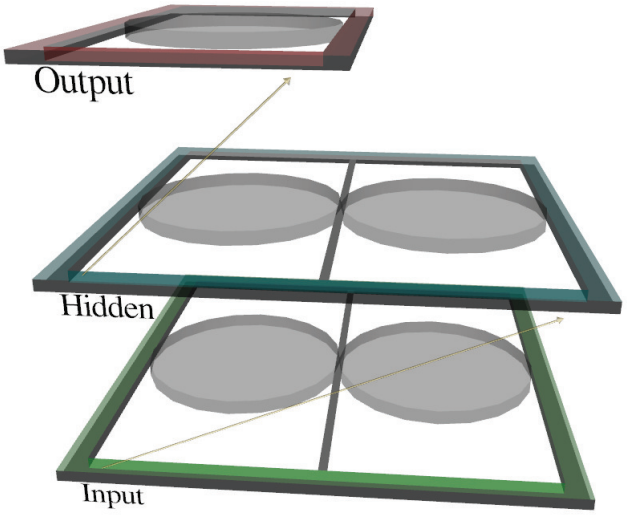

Figure 7: Standard network configuration with *Emergent*

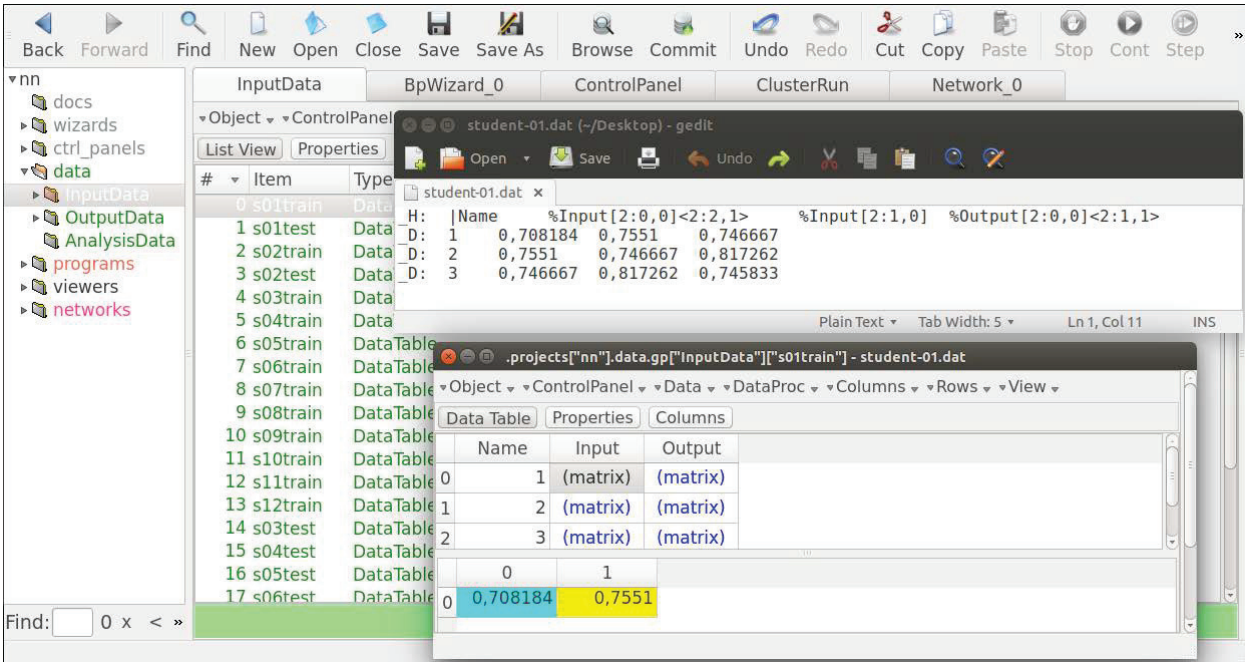

Figure 8: Network *StdInputData* and text files

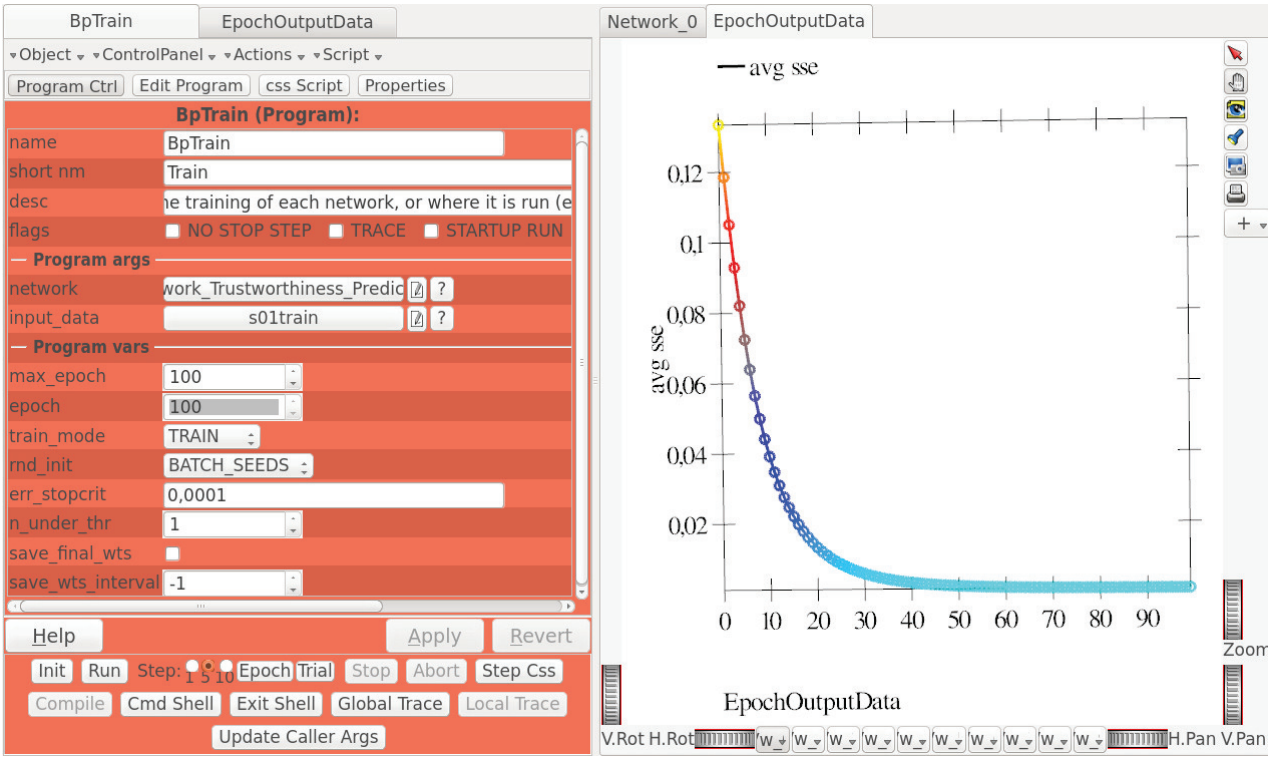

Figure 9: Training process parameters and simulation

Interestingly, regarding overall error prediction, the results reveal a notable similarity between the test and predicted values. However, the observed difference between the trustworthiness levels through the modules is not significant. Therefore, the model is suitable for this students' trustworthiness behaviour, but we cannot demonstrate the stability of this prediction approach for other cases (i.e. more differences in trustworthiness evolution).

With respect to e-assessment security, the most significant finding is related to detect anomalous user assessment. From these data, 2 students (student 6 and 9), whose error prediction is greater than 3%, were found anomalous and required further investigation for potential cheating in order to validate the authenticity of the students' learning process.

Finally, we discovered that the number of modules in the course (i.e. the slots or the points in time) must be increased. If the number of training instances is increased, the student's NN will be able to accurately predict more trustworthiness different cases (not only those cases with low variation in trustworthiness evolution).

6. Conclusions and Further Work

In this paper, we first motivated the need to improve information security in mobile online collaborative learning and in particular MCSCL supported by mobile devices. To this end, we justified the feasibility of an approach focused on functional solutions, namely, based on trustworthiness assessment and prediction. The study reviewed the main works in the literature on security in mobile collaborative learning, how trustworthiness assessment and prediction are related to security, the time factor in trustworthiness modelling, and trustworthiness existing methodologies.

Then, we proposed an innovative trustworthiness and security methodological approach to build secure collaborative activities devoted to offering a comprehensive guideline for e-Learning designers and managers. The architecture of the methodology is based on building trustworthiness learning components, trustworthiness analysis and data processing, and trustworthiness assessment and prediction. We first described the main theoretical features of our methodological approach and then, the summary of its key phases is presented. Finally, we detailed each phase by analysing the processes, data, and components involved in the methodology.

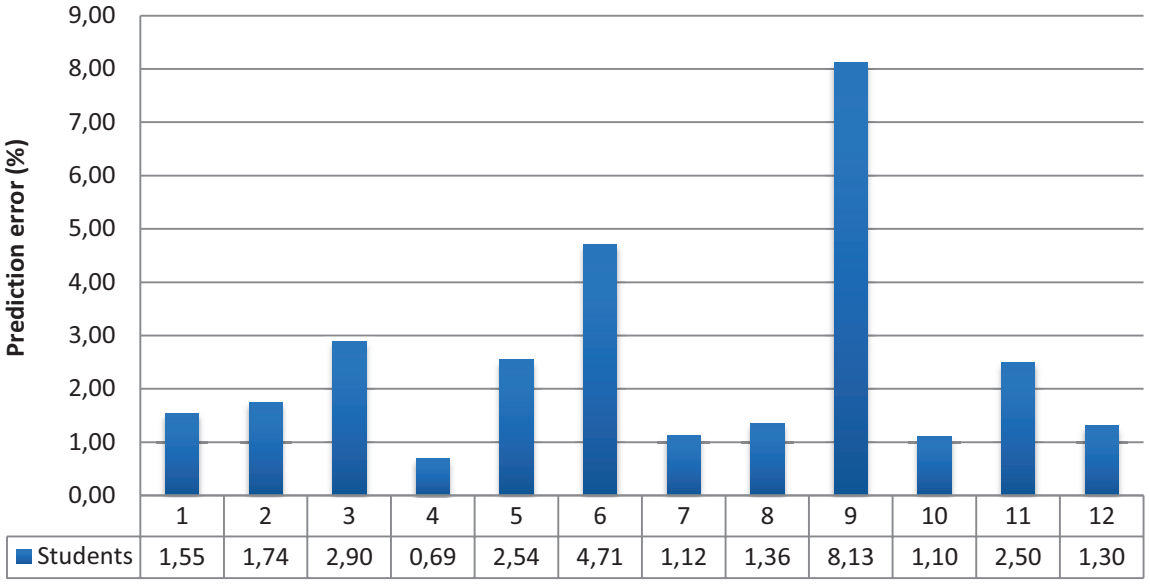

Figure 10: Students NN Prediction Results

The methodology was evaluated by presenting specific methods and techniques applied to real online courses. We used two studies, based on real online courses at the Open University of Catalonia, to evaluate and support the application and deployment of our trustworthiness methodology. Several significant aspects of our methodology were considered in terms of specific methods and techniques through their application in these real online courses.

Finally, we have presented an innovative prediction approach for trustworthiness behaviour to enhance security in online assessment. This study showed how neural network methods may support e-assessment prediction. These e-assessment prediction methods were performed in a real online course based on peer-to-peer assessment processes and mobile online collaborative activities. The processes and learning activities involved in the course were encapsulated as continuous assessment component. Moreover, from this component, we presented the design of trustworthiness history sequences with the aim of designing a neural network e-assessment proposal.

The most relevant findings that emerge from the results presented in this paper are related to trustworthiness methodological applications and trustworthiness prediction models. Regarding the trustworthiness methodology proposed, we supported the application and deployment of the methodology in two real online courses. The learning activities performed in the course were designed following the theoretical features, phases, data, and processes of our methodological approach. With respect to trustworthiness prediction, we demonstrated the feasibility of our neural network prediction approach. Regarding the overall error prediction, the results revealed a notable similarity between the test and predicted values. From these results, we were able to detect anomalous user assessment. From these data, 2 students, whose error prediction is greater than 3%, were found anomalous and required further investigation.

As ongoing work, we plan to continue the methodology testing and evaluation process by deploying its components in additional real online courses. Due to further deployments will require large amount of data analysis, we will continue investigating parallel processing methods to manage trustworthiness factors, indicators, and levels. Moreover, we would also like to investigate the use of location-based information of mobile learners to our approach, with the aim of improving trustworthiness assessment and then, trustworthiness prediction.

Finally, we discovered that the number of training instances should be increased. Therefore, with the aim of enhancing the prediction model, we plan to modify the learning activity presented in this study in order to generate more training instances. Hence, the student's neural network will be able to accurately predict more trustworthiness different cases (not only those cases with low variation in trustworthiness evolution).

Acknowledgement

This research was partly funded by the Spanish Government through the following projects: TIN2011-27076-C03-02 "CO-PRIVACY"; CONSOLIDER INGENIO 2010 CSD2007-0 004 "ARES"; TIN2013-45303-P "ICT-FLAG" Enhancing ICT education through Formative assessment, Learning Analytics and Gamification; and by funds from the Spanish Ministry for Economy and Competitiveness (MINECO) and the European Union (FEDER funds) under grant COMMAS (ref. TIN2013-46181-C2-1-R).

References

[1] T. Koschmann, Paradigm Shifts and Instructional Technology, in: T. Koschmann (Ed.), CSCL: Theory and Practice of an Emerging Paradigm, Lawrence Erlbaum Associates, Mahwah, New Jersey, 1996, pp. 1–23.

[2] Z. Luo, T. Zhang, A Mobile Service Platform for Trustworthy E-Learning Service Provisioning, in: S. Caballé, F. Xhafa, T. Daradoumis, A. A. Juan, Z. Luo, T. Zhang (Eds.), Architectures for Distributed and Complex M-Learning Systems, IGI Global, 2009, pp. 108–122. URL http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-60566-882-6

[3] S. Caballé, F. Xhafa, L. Barolli, Using mobile devices to support online collaborative learning, Mob. Inf. Syst. 6 (1) (2010) 27–47. URL http://dl.acm.org/citation.cfm?id=1804707.1804710

[4] J. Miguel, S. Caballé, J. Prieto, Information Security in Support for Mobile Collaborative Learning, in: The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013), IEEE Computer Society, Taichung, Taiwan, 2013, pp. 379–384. doi:10.1109/CISIS.2013.69.

[5] E. R. Weippl, Security in E-Learning, in: H. Bidgoli (Ed.), Handbook of information security Vol. 1, Key concepts, infrastructure, standards and protocols., Vol. 1, Wiley, Hoboken, NJ, 2006, pp. 279–293.

[6] C. J. Eibl, Discussion of Information Security in E-Learning, Ph.D. thesis, Universität Siegen, Siegen, Germany (2010). URL http://dokumentix.ub.uni-siegen.de/opus/volltexte/2010/444/pdf/eibl.pdf

[7] M. J. Dark, Information assurance and security ethics in complex systems: interdisciplinary perspectives, Information Science Reference, Hershey, PA, 2011.

[8] N. H. Mohd Alwi, I.-S. Fan, Information Security Threats Analysis for E-Learning, in: M. D. Lytras, P. Ordonez De Pablos, D. Avison, J. Sipior, Q. Jin, W. Leal, L. Uden, M. Thomas, S. Cervai, D. Horner (Eds.), Technology Enhanced Learning. Quality of Teaching and Educational Reform, Vol. 73 of Communications in Computer and Information Science, Springer Berlin Heidelberg, 2010, pp. 285–291, 10.1007/978-3-642-13166-0_41.

[9] K. M. Apampa, Presence verification for summative e-assessments, Ph.D. thesis, University of Southampton, Southampton, England (2010).

[10] Y. Levy, M. Ramim, A Theoretical Approach For Biometrics Authentication of E-Exams, in: Chais Conference on Instructional Technologies Research, The Open University of Israel, Raanana, Israel, 2006, pp. 93–101.

[11] J. Miguel, S. Caballé, F. Xhafa, J. Prieto, Security in Online Assessments: Towards an Effective Trustworthiness Approach to Support e-Learning Teams, in: 28th International Conference on Advanced Information Networking and Applications (AINA 2014), IEEE Computer Society, Victoria, Canada, 2014, pp. 123–130. doi:10.1109/AINA.2014.106.

[12] J. Miguel, S. Caballé, F. Xhafa, J. Prieto, A Massive Data Processing Approach for Effective Trustworthiness in Online Learning Groups, Concurrency and Computation: Practice and Experiencedoi:10.1002/cpe.3396. URL http://doi.wiley.com/10.1002/cpe.3396

[13] J. Miguel, S. Caballé, F. Xhafa, J. Prieto, L. Barolli, Towards a Normalized Trustworthiness Approach to Enhance Security in On-line Assessment, in: Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2014), IEEE Computer Society, Birmingham, UK, 2014, pp. 147–154. doi:10.1109/CISIS.2014.22.

[14] J. Miguel, S. Caballé, J. Prieto, Providing Information Security to MOOC: Towards effective student authentication, in: 5-th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2013), IEEE Computer Society, Xian, China, 2013, pp. 289 – 292. doi:10.1109/INCoS.2013.52.

[15] J. D. Demott, A. Sotirov, J. Long, Gray Hat Hacking, Third Edition Reviews, 3rd Edition, McGraw-Hill Companies, New York, 2011.

[16] CSO Magazine, US Secret Service, Software Engineering Insistute CERT Program at Carnegie Mellon University, Deloitte, 2011 Cybersecurity Watch Survey, Tech. rep., CSO Magazine (2011).

[17] Internet Crime Complaint Center, 2013 Internet Crime Report, Tech. rep., Bureau of Justice Assistance (2014). URL http://www.ic3.gov/media/annualreports.aspx

[18] Y. Laouris, N. Eteokleous, We need an Educationally Relevant Definition of Mobile Learning, Proc mLearn Cape Town (June) (2005) 1–13.

[19] D. Gambetta, Can We Trust Trust?, in: Trust: Making and Breaking Cooperative Relations, Blackwell, 1988, pp. 213–237.

[20] Y. Liu, Y. Wu, A Survey on Trust and Trustworthy E-learning System, in: 2010 International Conference on Web Information Systems and Mining, IEEE, 2010, pp. 118–122. doi:10.1109/WISM.2010.62. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5662295

[21] M. Raza, F. K. Hussain, O. K. Hussain, Neural Network-Based Approach for Predicting Trust Values Based on Non-uniform Input in Mobile Applications, Comput. J. 55 (3) (2012) 347–378. doi:10.1093/comjnl/bxr104. URL http://dx.doi.org/10.1093/comjnl/bxr104

[22] S. P. Marsh, Formalising Trust as a Computational Concept, Ph.D. thesis, University of Stirling (1994).

[23] I. Ray, S. Chakraborty, A Vector Model of Trust for Developing Trustworthy Systems, in: D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, P. Samarati, P. Ryan, D. Gollmann, R. Molva (Eds.), Computer Security – ESORICS 2004, Vol. 3193, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 260–275.

[24] S. Msanjila, H. Afsarmanesh, Automating Trust Assessment for Configuration of Temporary Partnerships, in: A. Azevedo (Ed.), Innovation in Manufacturing Networks, Vol. 266 of IFIP – The International Federation for Information Processing, Springer US, 2008, pp. 95–104.

[25] Z. Zhai, W. Zhang, The Estimation of Trustworthy of Grid Services Based on Neural Network, JNW 5 (10) (2010) 1135–1142.
URL http://dblp.uni-trier.de/db/journals/jnw/jnw5.html#ZhaiZ10

[26] K. Konrad, G. Fuchs, J. Barthel, Trust and electronic commerce-more than a technical problem, in: Reliable Distributed Systems, 1999. Proceedings of the 18th IEEE Symposium on, 1999, pp. 360–365. doi:10.1109/RELDIS.1999.805124.

[27] W. Song, V. Phoha, X. Xu, An adaptive recommendation trust model in multiagent system, in: Intelligent Agent Technology, 2004. (IAT 2004). Proceedings. IEEE/WIC/ACM International Conference on, 2004, pp. 462–465. doi:10.1109/IAT.2004.1342996.

[28] A. J. Flanagin, M. J. Metzger, Trusting expert- versus user-generated ratings online: The role of information volume, valence, and consumer characteristics, Computers in Human Behavior 29 (4) (2013) 1626 – 1634. doi:http://dx.doi.org/10.1016/j.chb.2013.02.001.
URL http://www.sciencedirect.com/science/article/pii/S0747563213000575

[29] X. Liu, A. Datta, A Trust Prediction Approach Capturing Agents' Dynamic Behavior, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three, IJCAI'11, AAAI Press, Barcelona, Catalonia, Spain, 2011, pp. 2147–2152. doi:10.5591/978-1-57735-516-8/IJCAI11-358.
URL http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-358

[30] F. Hussain, O. Hussain, E. Chang, Trustworthiness Measurement Methodology (TMM) for Assessment Purposes, in: Computational Cybernetics, 2007. ICCC 2007. IEEE International Conference on, 2007, pp. 107–112. doi:10.1109/ICCCYB.2007.4402024.

[31] M. Carbone, M. Nielsen, V. Sassone, A Formal Model for Trust in Dynamic Networks, in: IN PROC. OF INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND FORMAL METHODS (SEFM'03, Society Press, 2003, pp. 54–63.

[32] M. Wojcik, J. Eloff, H. Venter, Trust Model Architecture: Defining Prejudice by Learning, in: S. Fischer-Hübner, S. Furnell, C. Lambrinoudakis (Eds.), Trust and Privacy in Digital Business, Vol. 4083 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, pp. 182–191.

[33] P. Laforcade, Towards a UML-based educational modeling language, in: Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on, 2005, pp. 855 – 859. doi:10.1109/ICALT.2005.288.

[34] J. Miguel, S. Caballé, F. Xhafa, J. Prieto, Security in Online Web Learning Assessment. Providing an Effective Trustworthiness Approach to Support e-Learning Teams, World Wide Web (2015) 1–22doi:10.1007/s11280-014-0320-2.

[35] B. Mobasher, R. Burke, R. Bhaumik, C. Williams, Toward Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness, ACM Trans. Internet Technol.doi:10.1145/1278366.1278372.
URL http://doi.acm.org/10.1145/1278366.1278372

[36] B. Aisa, B. Mingus, R. O'Reilly, The Emergent neural modeling system, Neural Networks 21 (8) (2008) 1146 – 1152. doi:http://dx.doi.org/10.1016/j.neunet.2008.06.016.
URL http://www.sciencedirect.com/science/article/pii/S0893608008001287

Chapter 3

Conclusions and Further Work

In this chapter, we present the results obtained in terms of thesis achievements (see Section 3.1) and the future directions of research are suggested in Section 3.2.

3.1 Thesis Achievements

Here, we show the work already done and the results obtained in terms achievements of the thesis' objectives formulated in Section 1.3 and reported in our scientific publications. To this end, for each thesis' objective we relate the work reported in the published contributions with the corresponding objective by justifying the coherence between the publications and the thesis' object of research. Moreover, for each contribution we summarize a series of conclusions based on the results obtained in the published research. Therefore, in this section we relate the objectives of the thesis presented in Section 1.3 to the thesis objectives.

O1. Define a security CSCL model

In Miguel et al. (2012a), we argued that current e-learning systems supporting on-line collaborative learning do not sufficiently meet essential security requirements and this limitation can have a strong influence in the collaborative learning processes. In order to alleviate these problems we proposed an approach based on Public Key Infrastructure (PKI) models that offer services which ensure essential security properties in on-line

collaborative learning, such as availability, integrity, identification and authentication, access control, confidentiality, non repudiation, time stamping, audit service and failure control. Finally, we justified that these technological measures alleviate security problems, but they cannot reach a comprehensive security solution for CSCL.

In Miguel et al. (2012b), the problems caused in on-line collaborative learning processes by the lack of security are discussed and the main guidelines for the design of secure CSCL systems are proposed to guide developers in this domain to incorporate security as an essential requirement. We presented a first approach for designing secure LMS systems, whose central educational purpose is the provision and support of CSCL processes and activities, while security is the main and transversal requirement guiding the whole design process. To this end, we first provided background work in the context of LMS systems, CSCL, and main security services, such as PKI. Then, we provided general security services and requirements for e-Learning and specifically for CSCL. These approaches were finally merged to propose the main guidelines to design and develop Secure Collaborative Learning Management Systems (SCLMS) that focus on the specific support for CSCL with security as a guiding requirement to conduct the whole design process. Through the exploitation of SCLMS in real contexts, it was proposed to enhance and improve the CSCL experience of all participants of the collaboration.

In Miguel et al. (2013a), an overview of secure LMSs was presented, inspecting which the most relevant factors to consider are, and connecting this approach to specific aspects for mobile collaborative learning. Then, real-life experiences in security attacks in mobile learning were reported showing a practical perspective of the learning management system vulnerabilities. From this experience and considerations, the main guidelines for the design of security solutions applied to improve mobile collaborative learning were proposed. Therefore, this paper proposed a first approach for designing secure mobile collaborative LMS. To this end, we proposed a model based on security properties, attacks and PKI solutions and considering aspects related to CSCL and m-Learning.

In Miguel et al. (2013b), the lack of provision of IS to MOOC is investigated, with regards to anomalous user authentication, which cannot verify the actual student's identity to meet grading requirements as well as satisfy accrediting institutions. In order to overcome this issue, it is proposed a global user authentication model called MOOC Smart Identity Agent (MOOC-SIA). This model includes an innovative authentication

process aiming at overcoming the anomalous student authentication issue in MOOCs, which is considered one of the most important barriers currently found in the MOOC arena. Further, our model considered the massive feature of MOOCs and provides an authentication system flexible enough to meet each and every type of course and student profile. To this end, the MOOC-SIA model was designed to combine and set several authentication methods in MOOC platforms with the aim to offer a multi-fold solution combining different technologies, requirements and user resources. To meet most of these requirements, we included a modular PKI-based security model as the main component that manages different authentication methods in the MOOC platform in a centralized fashion. Moreover, an innovative use of data mining techniques for education was considered in order to infer potential anomalous authentication from tracking MOOC participants.

O2. Trustworthiness methodology

In Miguel et al. (2014c), we proposed a methodological approach to modelling trustworthiness in on-line collaborative learning. This proposal aims at building a theoretical approach to provide e-Learning designers and managers with guidelines for incorporating security into on-line collaborative activities through trustworthiness assessment and prediction. In this paper, we first motivated the need to improve information security in on-line collaborative learning with trustworthiness solutions. Then, we proposed an innovative trustworthiness and security methodological approach to build secure CSCL activities and devoted to offer a comprehensive guideline for e-Learning designers and managers. Finally, the methodology was evaluated by presenting specific methods and techniques applied to real on-line courses.

In Miguel et al. (2014b), we proposed a functional security model based on trustworthiness and collective intelligence. Both of these topics are closely related to on-line collaborative learning and on-line assessment models. Therefore, the main goal of this paper was to discover how security can be enhanced with trustworthiness in an on-line collaborative learning scenario through the study of the collective intelligence processes that occur on on-line assessment activities. To this end, a peer-to-peer public student's profile model, based on trustworthiness is proposed, and the main collective intelligence processes involved in the collaborative on-line assessments activities, were presented.

O3. Trustworthiness assessment and prediction

In Miguel et al. (2014e), an approach to enhance information security in on-line assessment based on a normalized trustworthiness model is presented. In this paper, it is justified why trustworthiness normalization is needed and a normalized trustworthiness model is proposed by reviewing existing normalization procedures for trustworthiness applied to e-assessments. To this end, we first motivated the need to improve information security in e-Learning and in particular in e-assessment. Then, we showed the feasibility of building security hybrid models, based on trustworthiness approaches. However, trustworthiness analysis in e-Learning requires normalization processes in order to tackle several trustworthiness modelling problems presented in the paper. As a main contribution of this paper, we proposed a methodological approach to build a normalized trustworthiness model. Finally, we used a real on-line course intended to evaluate a hybrid evaluation system supported by our normalized trustworthiness model.The experimental results showed the feasibility of modelling security by analysing normalized trustworthiness levels and indicators. From the results comparing manual evaluation and trustworthiness levels, it was inferred that it is viable to enhance security in e-assessment by modelling and normalizing trustworthiness behaviours.

In Miguel et al. (2014d), previous trustworthiness models were endowed with prediction features by composing trustworthiness modelling and assessment, normalization methods, history sequences, and neural network-based approaches. In order to validate our approach, a peer-to-peer e-assessment model was presented and carried out in a real on-line course. In particular, we presented an innovative prediction approach for trustworthiness behaviour to enhance security in on-line assessment and this study showed how neural network methods may support e-assessment prediction. To this end, we first motivated the need to improve information security in on-line assessment with trustworthiness solutions based on time factor models in order to analyse prediction techniques. Then, we conducted our research to peer-to-peer e-assessment, and trustworthiness sequences with the aim to apply the concept of trustworthiness sequences customized for the peer-to-peer e-assessment. Finally, we endowed our trustworthiness model with the prediction features by composing trustworthiness models, normalization, history sequences, and neural network models.

O4. Design secure CSCL activities and e-assessment

In Miguel et al. (2014a), a trustworthiness model for the design of secure collaborative learning e-assessment in CSCL is proposed is proposed. In this paper, we presented an innovative approach for modelling trustworthiness in the context of secure learning assessment in on-line collaborative learning groups. The study showed the need to propose a hybrid assessment model which combined technological security solutions and functional trustworthiness measures. This approach was based on trustworthiness factors, indicators and levels which allowed us to discover how trustworthiness evolves into the learning system. We proposed several research instruments for collecting students' data. Since extracting and structuring LMS data is a costly process, a parallel processing approach was proposed, which was fully developed and tested in Miguel et al. (2015b).

In Miguel et al. (2015b), a trustworthiness-based approach for the design of secure learning activities in on-line learning groups was proposed by the presentation of guidelines of a holistic holistic security model in on-line collaborative learning through an effective trustworthiness approach. To this end, we first motivated the need to improve information security in e-Learning and in particular in CSCL activities. Then, we proposed a methodological approach to build a security model for CSCL activities with the aim to enhance standard technological security solutions with trustworthiness factors and rules. As a result, the guidelines of a holistic security model in on-line collaborative learning through an effective trustworthiness approach were first proposed. However, as learners' trustworthiness analysis involves dealing with large amount of data generated along learning activities, processing this information is computationally costly, especially if required in real-time. To this end, and as a main contribution of this paper, we developed and tested a parallel processing approach that can considerably decrease the time of data processing, thus allowing for building relevant trustworthiness models to support learning activities even in real-time.

O5. Experimentation and validation

Finally, in the last set of results, we combine all the objectives. In short:

O1 Define a security CSCL model.

O2 Trustworthiness methodology.

O3 Trustworthiness assessment and prediction.

O4 Design secure CSCL activities and e-assessment.

O5 Experimentation and validation.

We have carried out three studies in our real context of e-Learning of the UOC during academic term of 2014 and 2015, with the aim to experiment with specific trustworthiness and security approaches devoted to evaluate the feasibility of our trustworthiness models, tools, and methodologies (Miguel et al., 2015a,b,c). These studies are presented in the rest of this section and the key features of the pilots can be summarized as follows:

In the first study (Miguel et al., 2014e) students' evaluation was based on a hybrid continuous assessment model by using several manual and automatic assessment instruments. There were 12 students distributed in three groups and the course was arranged in four stages. These stages were taken as time references in order to implement trustworthiness sequences. At the end of each collaborative stage, each student had to complete a survey. The coordinator of the group had to complete two reports, public and private, and at the end of each stage, the members the group was evaluated by the coordinator. General e-Learning activities were supported by UOC Virtual Campus web-based services, whilst the collaborative activities were supported by the UOC's BSCW. BSCW is a fully web-based collaboration platform that facilitates efficient teamwork through a wide range of functions (OrbiTeam, 2015). The UOC's BSCW offered both rating systems and general learning management indicators.

The second study (Miguel et al., 2014d) extended the scope of the first one to a more standard scenario. In this context, we were able to apply massive deployment for automatic e-assessment processes. The course was focused on peer-to-peer e-assessment and it has the following main features:

- 12 students performed a subjective peer-to-peer e-assessment, that is, each student was able to assess the rest of class peers in terms of knowledge acquired and participation in the class assignments.

- The course followed seven stages which were taken as time references in trustworthiness analysis. These time references allow us to compare trustworthiness evolution as well as to carry out e-assessment methods.
- Each stage corresponded to a module of the course, which had a learning component (i.e. book) that the student should have studied before developing the assessment activities of the course.
- Students' e-assessment was based on a manual continuous e-assessment model by using several manual e-assessment instruments.
- Manual e-assessment was complemented with automatic methods, which represented up to 20% of the total student's overall course grade.
- Taking into account the previous features, we implemented a hybrid e-assessment method by combining manual and automatic e-assessment methods, and the model allowed us to compare results in both cases.

Finally the third study (Miguel et al., 2015c) followed the same design model than the first one. However, we incorporated the enhancements detected during the first development. We denote this courses as follow:

- *CSCL-course-1*
  Hybrid assessment based on collaborative activities.
- *p2p-course*
  Automatic peer-to-peer e-assessment processes.
- *CSCL-course-2*
  The second development of *CSCL-course-1* (the same design model than *CSCL-course-1*).

Moreover, the main thesis' outcomes related to experimental results are presented in the rest of this section following the next 4 topics:

1. Anomalous user assessment.
2. Peer-to-peer visualization tools.

3. Trustworthiness prediction achievements.

4. Massive data processing.

O5.1. Experimentation and validation: anomalous user assessment

In Miguel et al. (2015b), we proposed a trustworthiness model for the design of secure learning assessment in on-line web collaborative learning groups. In this paper, a holistic security model was designed, implemented and evaluated in a real context of e-Learning. Implications of this study were remarked for secure assessment in on-line collaborative learning through effective trustworthiness approaches. The study showed the need to propose a hybrid assessment model which combines technological security solutions and functional trustworthiness measures. To this end, a holistic security model was designed, implemented and evaluated in a real context of e-Learning. This approach is based on trustworthiness factors, indicators and levels, which allowed us to discover how trustworthiness evolves into the learning system.

From the trustworthiness methodology, we designed the peer-to-peer e-assessment component. The overall assessment model was a hybrid (i.e. automatic and manual) approach formed by:

- The peer-to-peer e-assessment component, which is automatic and the students were assessed by students.
- The continuous assessment activities, which is manual and the tutors assessed the students.

The results reveal a notable difference between the overall range of these values. Fig. 3.1 shows that most of peer-to-peer assessment values are in the range from 3,5 to 4,3 (the e-assessment scale was from 1 to 5) and the continuous assessment, from 1 to 9.

The statistical analysis showed significant findings regarding the feasibility of the hybrid evaluation method. The results of the comparisons between manual and automatic e-assessment indicate (also see Fig. 3.1):

- The mean difference between manual and automatic method is 0,81 (the scale used from 0 to 10).

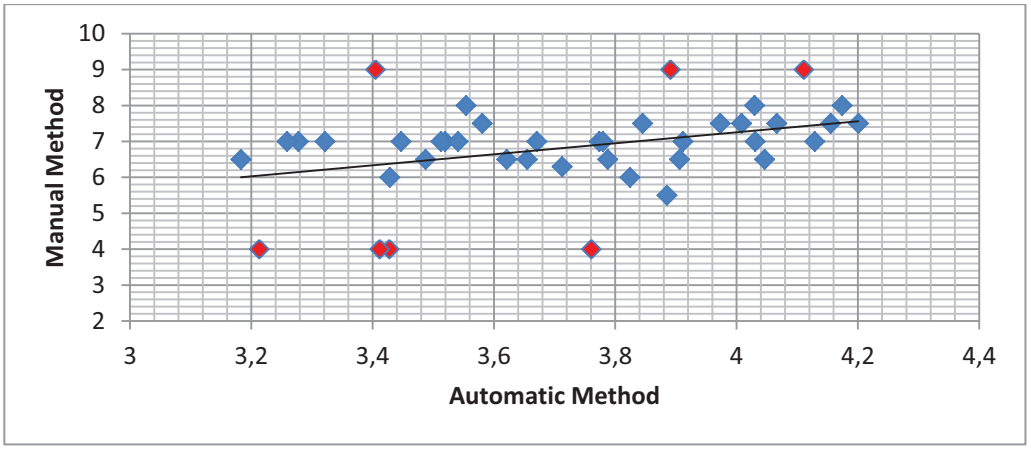

FIGURE 3.1: Manual and automatic evaluation methods (dispersion chart)

- The maximum and minimum difference: 0,03 and 2,82.
- The percentage of assessment cases in which the difference between manual and automatic assessment is less than 1 (i.e. 10% with respect the maximum score) is the 76,92%.
- If we extend the difference to more than 2 points in the scale, the percentage of assessment cases in this range is the 92,31%.

The most significant finding was related to anomalous user assessment. From these data, 3 students whose deviation was greater than 20% were found anomalous and required further investigation for potential cheating in order to validate the authenticity (i.e. identification and integrity) of her learning processes and results.

O5.2. Experimentation and validation: peer-to-peer visualization tools

In Miguel et al. (2015a), we presented a peer-to-peer on-line assessment approach carried out in a real on-line course developed in our real e-Learning context of the Open University of Catalonia. The design presented in this paper was conducted by our trustworthiness security methodology with the aim of building peer-to-peer collaborative activities, which enhances security e-Learning requirements.

Peer-to-peer visualizations methods were proposed to manage security e-Learning events, as well as on-line visualization through peer-to-peer tools, intended to analyse collaborative relationship for monitoring the collaborative learning process by tutors and students.

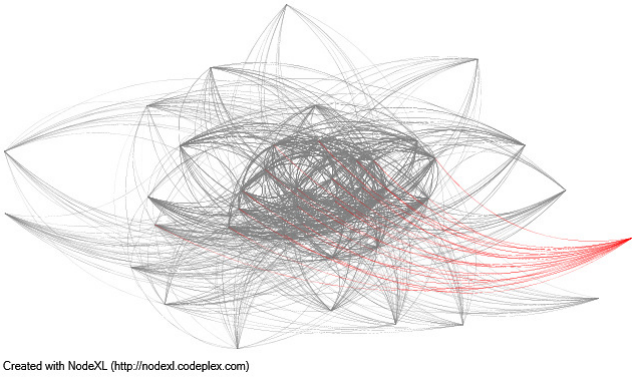

FIGURE 3.2: The students' e-assessment relationships

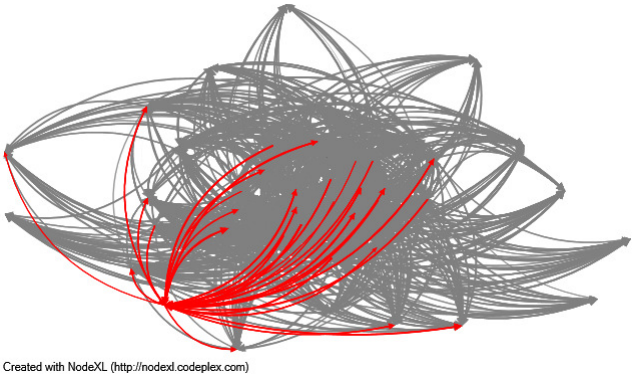

FIGURE 3.3: Weighted students' e-assessment relationships

As shown in Fig. 3.2, the tutor can select a student and the vertex for the students' e-assessment relationships are remarked. The remarked edges correspond to the assessment relation between a student in the peer-to-peer process, that is, those students who assessed the selected student and the students who were assessed by the selected student. We also introduced the score value assessed by each student by using an edge weight column. For this reason, the score value is also represented as a edge weight in the graph (see Fig. 3.3).

O5.3. Experimentation and validation: trustworthiness prediction achievements

In Miguel et al. (2015c), we proposed a methodological approach to modelling trustworthiness in on-line collaborative learning. Our proposal sets out to build a theoretical approach with the aim to provide e-Learning designers and managers with guidelines for incorporating security into on-line collaborative activities through trustworthiness assessment and prediction. We proposed an innovative trustworthiness and security

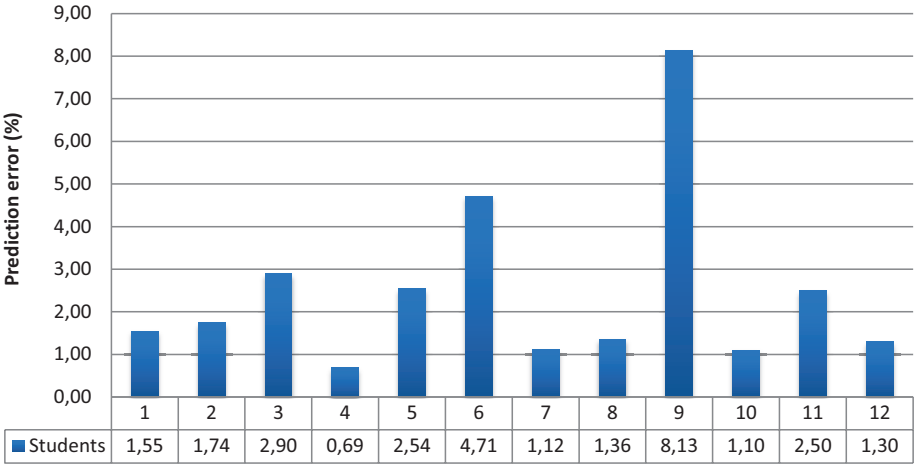

FIGURE 3.4: Students NN prediction results

methodological approach to build secure collaborative activities devoted to offering a comprehensive guideline for e-Learning designers and managers. The architecture of the methodology was based on building trustworthiness learning components, trustworthiness analysis and data processing, and trustworthiness assessment and prediction. The methodology was evaluated by presenting specific methods and techniques applied to real on-line courses.

We used two studies (*CSCL-course-1* and *p2p-course*), based on real on-line courses at the Open University of Catalonia, to evaluate and support the application and deployment of our trustworthiness methodology. Several significant aspects of our methodology were considered in terms of specific methods and techniques through their application in these real on-line courses. In the same study, we presented an innovative prediction approach for trustworthiness behaviour to enhance security in on-line assessment. This study showed how neural network methods may support e-assessment prediction. These e-assessment prediction methods were performed in a real on-line course based on peer-to-peer assessment processes and mobile on-line collaborative activities. The processes and learning activities involved in the course, were encapsulated as continuous assessment component. Moreover, from this component, we presented the design of trustworthiness history sequences with the aim of designing a neural network e-assessment proposal.

The most relevant findings that emerge from the results presented in Miguel et al. (2015), were related to trustworthiness methodological applications and trustworthiness prediction models. Regarding the trustworthiness methodology proposed, we supported

the application and deployment of the methodology in the real on-line courses presented ((*CSCL-course-1* and *p2p-course*)). The learning activities performed in the course were designed following the theoretical features, phases, data, and processes of our methodological approach.

With respect to trustworthiness prediction we focused the application of trustworthiness prediction on the course *p2p-course*. We demonstrated the feasibility of our neural network prediction approach. Regarding the overall error prediction, the results revealed a notable similarity between the test and predicted values. From these results (see Fig. 3.4), we were able to detect anomalous user assessment. From these data, 2 students, whose error prediction is greater than 3%, were found anomalous and required further investigation.

O5.4. Experimentation and validation: massive data processing

In Miguel et al. (2015c), the implementation of our parallel approach faced two important challenges: handle several formats of logs files coming from different LMS and the large size of these log files. We showed how to normalize different log file structures as an input for the MapReduce paradigm to manage huge amounts of log data in order to extract the trustworthiness information defined in our model.

We used distributed infrastructure, Hadoop and Cluster Computing, to implement and evaluate our parallelization approach for massive processing of log data. Experimental results showed the feasibility of coping with the problem of structuring and processing ill-formatted, heterogeneous, large log files to extract information on trustworthiness indicators and levels from learning groups and ultimately fill a global framework devoted to improve information security in e-Learning in real-time. We eventually conclude that it is viable to enhance security in CSCL activities by our trustworthiness model, though taking on the overhead caused by the use of distributed infrastructure for massive data processing.

Fig. 3.5 shows comparative results of the battery of tests with multiple Hadoop nodes (i.e. 2, 4, 6, 8 and 10 workers). Note that 0-node shows results of local sequential processing benchmark. From this experimental study, we can see that the results did not grow linearly anymore. By using a distributed MapReduce Hadoop infrastructure,

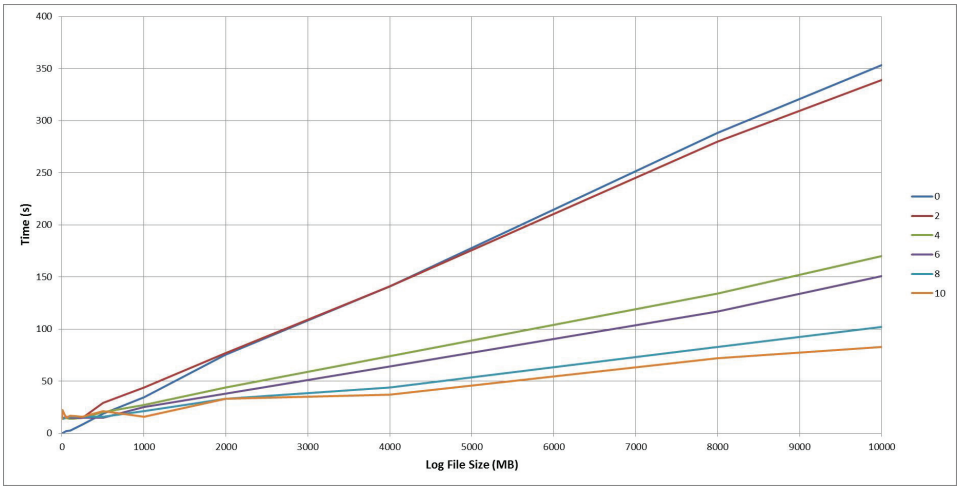

FIGURE 3.5: Comparative map reduce results

a considerable speed up is achieved in processing large log file data as shown in Fig. 3.5. Regarding log file size, on the one hand, for too small values, the overhead introduced by the MapReduce framework when sending the parts to the nodes and combining output data, is noticeable. Also, the framework control tasks spends too much time managing and distributing data. On the other hand, values of the task size close to 3,000 MB considerably diminished this amount of time in comparison with the total processing time.

3.2 Further Work

As ongoing work, the methodology testing and evaluation can be continued by deploying collaborative learning e-assessment components in additional real on-line courses. Although the first pilot was developed in two academic terms, we detected potential improvements for the second on-line course. For this reason, we plan to apply these improvements in the next academic term.

Due to further deployments will require a large amount of data analysis, we propose to enhance the parallel processing methods proposed in this research with the aim to manage trustworthiness factors and indicators. These enhancements can be reached through improving the MapReduce configuration strategies that would result in improvement of a parallel speed-up, such as customized size of partitions.

Moreover, the next steps in trustworthiness predictions methods will focus on evaluation and test processes. With respect to our prediction approach, in order to predict both trustworthiness students' behaviour and evaluation alerts (e.g. anomalous results), we suggest to evaluate additional neural networks approaches and data mining models.

The next steps regarding the prediction model will go through investigating the use of location-based information of mobile learners to our approach, with the purpose of improving trustworthiness assessment and then, trustworthiness prediction. In addition, we discovered that the number of training instances should be increased. Therefore, with the aim of enhancing the prediction model, we suggest to modify the learning activity presented in this study in order to generate more training instances. Hence, the student's neural network will be able to more accurately predict more trustworthiness different cases (not only those cases with low variation in trustworthiness evolution).

In addition, ongoing work will be implementing e-assessment solutions based on the proposed methodology that will extend the set of security properties considered as requirements, such as, privacy, non-repudiation, and so on. (i.e. not only identity and integrity, that were selected as the most significant security properties in the scope of e-assessment processes). To this end, we can analyse e-Learning cases and scenarios in which these properties become relevant requirement. For instance, in Miguel et al. (2014b) we defined privacy trustworthiness fields devoted to publish students' information according to the regulations, responsiveness, and the protection of privacy principles established for each educational institution.

Regarding visualization methods in peer-to-peer e-assessment models, we plan to enhance the proposed visualization tools by conducting additional testing activities. The visualizations capabilities for the tutors can be completed with additional facilities, specially those capabilities related to cope with anomalous user identification and integrity.

Although we have tackled the problem of predicting trustworthiness with neural network approaches, there exist other trustworthiness models without neural networks methods, such as similarity approaches and collaborative filtering (Flanagin and Metzger, 2013; Liu and Datta, 2011). Collaborative filtering is a process of finding similar users, computing predicted ratings, and applying the predictions, such as recommendations to the user (Soboroff and Nicholas, 2000). We will consider these approaches in order to predict

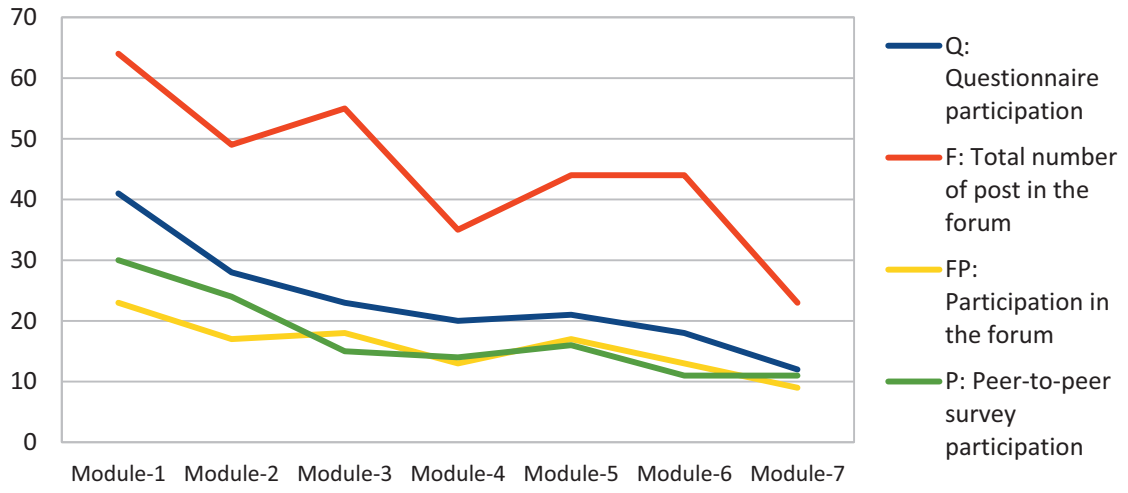

FIGURE 3.6: Students' participation evolution

students' trustworthiness. In this case, the item involved in the recommendation system could be the students themselves, in other words, the collaborative filtered system would be formed by students assessing other students. Then, the recommendation target would be the student's trustworthiness level. From this view, when a collaborative filtering system generates predictions for a target user, the system first identifies the other users whose interests correlate to the target user (i.e. user's neighbours). Hence, we will consider that if a tutor performs manual assessment, those students, which are tutor' neighbours, would be trustworthy students. These statements have not been applied and validated in the context of trustworthiness prediction in e-assessment, thus we propose to tackle the problem of predicting trustworthiness with a innovative collaborative filtering approach which will be validated in our real e-Learning context.

Regarding the pilots developed in real on-line courses, the participation level experimented a notable decrease along the course, especially at the end of e-assessment activities, due to the students' peak workload in their courses (see Fig. 3.6). As ongoing work, we plan to cope with alternative course schedule, with the aim to balance the students' peer-to-peer activities and other students' assignments.

Given the low number of students in the pilots, we were able to manually analyse the data in much more detail and flexibility. Likewise, we could experiment with several design alternatives adapting the model to the design cycles proposed in the methodology.

However, we have not extended the application of the methodology to larger scales. Therefore, we plan to propose a real on-line e-assessment model in a massive learning activity, such as in the contexts of MOOCs.

Regarding the student profile proposed, we plan to evaluate and test our students' profile model in real on-line courses. Due to these deployments will require large amount of data analysis, we will analyse and monitor the critical mass related to the students community in a sustainable fashion.

Trustworthiness analysis and the application of a trustworthiness methodology are topics that extend the scope of e-assessment and CSCL. In this context, we plan to consider new applications where trustworthiness assessment and prediction are suitable methods involved in security processes. In particular, we also plan to address our future research to social networks and how to enhance security requirements in this context based on trustworthiness solutions.

Moreover, we are aware of the special relevance that technological security solutions, such as PKI or biometry, had represented in many IT systems. Therefore, we also plan to investigate on hybrid models based on technological and functional approaches. We believe that trustworthiness and PKI-based approaches potentially form a comprehensive and solid security solution.

Finally, we are working on Elsevier book proposal with the aim of condensing our research work and achievements into a comprehensive and self-contained reference, in order to enhance security in CSCL through functional approaches based on trustworthiness.

Appendix A

Programme committees

This appendix includes 5 memberships to program committees of international conferences and workshops in the scope of the thesis work that the candidate has joined. This implied for the candidate to perform relevant research tasks, such as paper reviewer.

A.1 Technical Committee of CISIS 2015

Ninth International Conference on Complex, Intelligent, and Software Intensive Systems

Track 4: E-Learning and Groupware Systems

Chairs:

Santi Caballe, Open University of Catalonia, Spain
Yoshinari [illegible], Okayama University, Japan
Ravindra Dastikop, SDM College of Engineering and Technolgy, India

PC Members:

Thanasis Daradoumis , University of Aegean , GR
Jordi Conesa , Open University of Catalonia , ES
Michael Feidakis , University of Aegean , GR
Nestor Mora , Open University of Catalonia , ES
Jorge Moneo , Open University of Catalonia , ES
David Ganan , Open University of Catalonia , ES
Luis Casillas , University of Guadalajara , MX
Kaoru Sugita , Fukuoka Institute of Technology , JP
Yoshiaki Kasahara , Kyushu University , JP
Shunsuke Mihara , Lockon Inc. , JP
Shunsuke Oshima , Kumamoto National College of Technology , JP
Yuuichi Teranishi , NICT , JP
Kazunori Ueda, Kochi University of Technology, JP

A.2 Technical Committee of ALICE 2015

Fifth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional approaches.

PROGRAM COMMITTEE

Jordi Conesa, Open University of Catalonia, Spain
Thanasis Daradoumis, University of the Aegean, Greece
Michalis Feidakis, University of the Aegean, Greece
Angelo Gaeta, University of Salerno, Italy
David Gañán, Open University of Catalonia, Spain
Giuseppe Guarino, University of Salerno, Italy
Giuseppina Rita Mangione, Institute of Educational Documentation, Innovation and Research, Italy
Jorge Miguel, Open University of Catalonia, Spain
Néstor Mora, Open University of Catalonia, Spain
Anna Pierri, University of Salerno, Italy

A.3 Technical Committee of CISIS 2014

Eight International Conference on Complex, Intelligent, and Software Intensive Systems.

Track 17: Collaborative Learning using Social Networks

Chairs:

Santi Caballe, Open Un[illegible]talonia, Spain
David Britch, CM Group, UK

PC Members:

Alex Mackman, CM Group, UK
John Devaney, CM Group, UK
Steve Cox, CM Group, UK
Graham Papworth, CM Group, UK
Darren Dancey, Manchester Metropolitan University, UK
Thanasis Daradoumis, University of Aegean, GR
Jordi Conesa, Open University of Catalonia, ES
Michael Feidakis, University of Aegean, GR
Roxana Bassi, Open University of Catalonia, ES
Nestor Mora, Open University of Catalonia, ES
Jorge Moneo, Open University of Catalonia, ES
David Gañán, Open University of Catalonia, ES
Nicola Capuano, University of Salerno, IT
Jose Mangione, University of Salerno, IT
Anna Pierri, University of Salerno, IT
Luis Casillas, University of Guadalajara, MX

A.4 Technical Committee of INCoS 2014

Sixth International Conference on Intelligent Networking and Collaborative Systems.

-Track 9: SOFTWARE ENGINEERING, SEMANTICS AND ONTOLOGIES FOR INTELLIGENT NETWORKING AND COLLABORATIVE SYSTEMS

PC Members:

Nestor Mora, Open University of Catalonia, Spain
Jorge Moneo, Open University of Catalonia, Spain
David Ganan , Open University of Catalonia, Spain
Robert Clariso, Open University of Catalonia, Spain
Jordi Casas, Open University of Catalonia, Spain
Macario Polo, University of Castile-La Mancha, Spain
David Baneres, Open University of Catalonia, Spain
Elena Planas, Open University of Catalonia, Spain
Antoni Perez, Open University of Catalonia, Spain
Xiaodong Liu, Edinburgh Napier University, UK
Ricardo Martin, University of Valladolid, Spain
Elena Rodriguez, Open University of Catalonia, Spain
Isabel Guitart, Open University of Catalonia, Spain

A.5 Technical Committee of ALICE 2014

Fourth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional approaches.

PROGRAM COMMITTEE

Jordi Conesa, Open University of Catalonia, Spain
Luis Casillas, University of Guadalajara, Mexico
Thanasis Daradoumis, University of the Aegean, Greece
Stavros Demetriadis, Aristotle University of Thessaloniki, Greece
Ian Dunwell, Serious Games Institute, UK
Michalis Feidakis, University of the Aegean, Greece
Angelo Gaeta, University of Salerno, Italy
David Gañán, Open University of Catalonia, Spain
Giuseppina Rita Mangione, University of Salerno, Italy
Jorge Miguel, Open University of Catalonia, Spain
Néstor Mora, Open University of Catalonia, Spain
Anna Pierri, University of Salerno, Italy
Roberto Pirrone, University of Palermo, Italy
Pier Giuseppe Rossi, University of Macerata, Italy

Bibliography

Ackland, R., Hansen, D. L., Shneiderman, B., and Smith, M. A. (2011). *Analyzing social media networks with NodeXL: insights from a connected world*. Elsevier, Morgan Kaufmann, Amsterdam [u.a.

Akker, J. (1999). Principles and methods of development research. In Akker, J., Branch, R., Gustafson, K., Nieveen, N., and Plomp, T., editors, *Design Approaches and Tools in Education and Training*, pages 1–14. Springer Netherlands.

American Psychological Association (2010). *Publication manual of the American Psychological Association*. American Psychological Association, Washington, DC, 6th ed edition.

Apampa, K. M. (2010). *Presence verification for summative e-assessments*. PhD thesis, University of Southampton, Southampton, England.

Bazán, R., Barrio, J. A., and **Miguel, J.** (2010). Servicios TIC en universidad san jorge: un modelo eficiente en infraestructuras de sistemas y comunicaciones. Zaragoza, Spain.

Bazán, R. and **Miguel, J.** (2009). Entorno de movilidad seguro en la universidad san jorge. Feria de Madrid (IFEMA), Madrid, Spain.

Bernthal, P. (1997). A survey of trust in the workplace. Executive summary, HR Benchmark Group, Pittsburg, PA.

Besimi, A., Shehu, V., Abazi-Bexheti, L., and Dika, Z. (2009). Managing security in a new learning management system (LMS). pages 337–342, Cavtat, Croatia. IEEE Computer Society.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4–5):175 – 308. doi:http://dx.doi.org/10.1016/j.physrep.2005.10.009.

Caballé, S. (2008). Combining generic programming and service-oriented architectures for the effective and timely development of complex e-learning systems. pages 94–100, Barcelona, Spain. IEEE Computer Society.

Caballé, S. and Feldman, J. (2008). CoLPE: Communities of learning practice environment. In Foster, D. and Schuler, D., editors, *Proceedings of the Directions and Implications of Advanced Computing*, Berkeley, CA, USA.

Carullo, G., Castiglione, A., Cattaneo, G., Santis, A. D., Fiore, U., and Palmieri, F. (2013). FeelTrust: Providing trustworthy communications in ubiquitous mobile environment. *2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, 0:1113–1120. doi:http://doi.ieeecomputersociety.org/10.1109/AINA.2013.100.

Cheswick, W. R., Bellovin, S. M., and Rubin, A. D. (2003). *Firewalls and Internet security : repelling the wily hacker*. Addison-Wesley, Boston.

CSO Magazine, US Secret Service, Software Engineering Insistute CERT Program at Carnegie Mellon University, and Deloitte (2011). 2011 cybersecurity watch survey. Technical report, CSO Magazine.

Dai, C., Lin, D., Bertino, E., and Kantarcioglu, M. (2008). An approach to evaluate data trustworthiness based on data provenance. In Jonker, W. and Petković, M., editors, *Secure Data Management*, volume 5159, pages 82–98. Springer Berlin Heidelberg, Berlin, Heidelberg.

Daradoumis, T., Martínez-Monés, A., and Xhafa, F. (2006). A layered framework for evaluating on-line collaborative learning interactions. *International Journal of Human-Computer Studies*, 64(7):622 – 635. doi:http://dx.doi.org/10.1016/j.ijhcs.2006.02.001.

Dark, M. J. (2011). *Information assurance and security ethics in complex systems: interdisciplinary perspectives*. Information Science Reference, Hershey, PA.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In Dillenbourg, P., editor, *Collaborative-learning: Cognitive and Computational Approaches*, pages 1–19. Elsevier Science, Oxford, UK.

Dillenbourg, P., editor (2003). *Collaborative learning: cognitive and computational approaches*. Advances in learning and instruction series. Elsevier, Amsterdam, 2 edition.

Eibl, C. J. (2010). *Discussion of Information Security in E-Learning*. PhD thesis, Universität Siegen, Siegen, Germany.

Equipo de Serguridad de RedIRIS (2013). Informe de incidentes de seguridad año 2012. Technical report, Red Académica y de Investigación Española (RedIRIS).

Erwin, T. D. (1992). Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes. *The Journal of Higher Education (JHE)*, 63(4):463–465.

Flanagin, A. J. and Metzger, M. J. (2013). Trusting expert- versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, 29(4):1626 – 1634. doi:http://dx.doi.org/10.1016/j.chb.2013.02.001.

Gambetta, D. (1988). Can we trust trust? In *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Blackwell.

Gil-Albarova, A., Martínez, A., Tunnicliffe, A., and **Miguel, J.** (2013). Estudiantes universitarios y calidad del plan de acción tutorial. valoraciones y mejoras. *REDU. Revista de Docencia Universitaria*, 11(2).

Harris, S. (2002). *All-In-One CISSP Certification Exam Guide*. McGraw-Hill Osborne Media, New York.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Q.*, 28(1):75–105.

Horak, Z., Kudelka, M., Snasel, V., Abraham, A., and Rezankova, H. (2011). Forcoa.NET: An interactive tool for exploring the significance of authorship networks in DBLP data. In *Computational Aspects of Social Networks (CASoN), 2011 International Conference on*, pages 261–266. 10.1109/CASON.2011.6085955.

Hussain, F., Hussain, O., and Chang, E. (2007). Trustworthiness measurement methodology (TMM) for assessment purposes. In *Computational Cybernetics, 2007. ICCC 2007. IEEE International Conference on*, pages 107–112. 10.1109/ICCCYB.2007.4402024.

Hussain, O., Chang, E., Hussain, F., and Dillon, T. (2009). Determining the failure level for risk analysis in an e-commerce interaction. In Dillon, T., Chang, E., Meersman, R., and Sycara, K., editors, *Advances in Web Semantics I*, volume 4891 of *Lecture Notes in Computer Science*, pages 290–323. Springer Berlin Heidelberg.

IETF (2011). Public-key infrastructure (x.509) (pkix) - documents.

Koschmann, T. (1996). Paradigm shifts and instructional technology. In Koschmann, T., editor, *CSCL: Theory and Practice of an Emerging Paradigm*, pages 1–23. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Kudelka, M., Horak, Z., Snasel, V., and Abraham, A. (2010). Social network reduction based on stability. In *Computational Aspects of Social Networks (CASoN), 2010 International Conference on*, pages 509–514. 10.1109/CASoN.2010.120.

Levy, Y. and Ramim, M. (2006). A theoretical approach for biometrics authentication of e-exams. In *Chais Conference on Instructional Technologies Research*, pages 93–101, The Open University of Israel, Raanana, Israel.

Liu, X. and Datta, A. (2011). A trust prediction approach capturing agents' dynamic behavior. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, IJCAI'11, pages 2147–2152, Barcelona, Catalonia, Spain. AAAI Press. 10.5591/978-1-57735-516-8/IJCAI11-358.

Liu, Y. and Wu, Y. (2010). A survey on trust and trustworthy e-learning system. In *2010 International Conference on Web Information Systems and Mining*, pages 118–122. IEEE. 10.1109/WISM.2010.62.

Marsh, S. P. (1994). *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling.

Miguel, J. (2005a). Desarrollo de aplicaciones web avanzadas utilizando software libre. Teruel, Spain.

Miguel, J. (2005b). Mesa redondas sobre software libre en la empresa. Zaragoza, Spain.

Miguel, J. (2006a). Aspectos de seguridad en los documentos firmados electrónicamente. Teruel, Spain.

Miguel, J. (2006b). El proyecto MoodlePKI. Tarragona, Spain.

Miguel, J. (2007a). Aplicaciones de las TIC en la formación. Parque Tecnológico Walqa, Huesca, Spain.

Miguel, J. (2007b). El proyecto SECURLEFIS. Badajoz, Spain.

Miguel, J. (2007c). Utilización de las nuevas tecnologías en la innovación educativa. Parque Tecnológico Walqa, Huesca, Spain.

Miguel, J. (2008). Software libre: el futuro de la informática. Zaragoza, Spain.

Miguel, J. (2011a). Educación superior: desafíos y estrategias de las TIC en un entorno de globalización. Madrid, Spain.

Miguel, J. (2011b). Seguridad en entornos virtuales de aprendizaje: Un ejemplo. In Cabezudo Rodríguez, N., editor, *Inclusión digital: perspectivas y experiencias*, pages 311–336. Prensas Universitarias de Zaragoza, Zaragoza.

Miguel, J. (2013a). E-learning y seguridad, diseño de entornos virtuales de aprendizaje seguros para modelos de e-learning actuales y nuevas tendencias. Universitat Oberta de Catalunya, Barcelona, Spain.

Miguel, J. (2013b). El DNI electrónico en universidad san jorge: Caso de uso y acciones previstas. Centro Demostrador del DNI electrónico en Aragón, Huesca, Spain.

Miguel, J. (2014). Flexilabs, el laboratorio para estudiantes disponible en cualquier momento y desde cualquier lugar. IFEMA, Madrid, Spain.

Miguel, J., Caballé, S., and Prieto, J. (2012a). Security in learning management systems: Designing collaborative learning activities in secure information systems. *eLearning Papers. European Comission*, 28:1–3. elearningeuropa.info. ISSN: 1887-1542.

Miguel, J., Caballé, S., and Prieto, J. (2012b). Providing security to computer-supported collaborative learning systems: An overview. In *Fourth IEEE International Conference on Intelligent Networking and Collaborative Systems (INCOS 2012)*, pages 97–104, Bucharest, Romania. IEEE Computer Society. 10.1109/iNCoS.2012.60.

Miguel, J., Caballé, S., and Prieto, J. (2013a). Information security in support for mobile collaborative learning. In *The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013)*, pages 379–384, Taichung, Taiwan. IEEE Computer Society. 10.1109/CISIS.2013.69.

Miguel, J., Caballé, S., and Prieto, J. (2013b). Providing information security to MOOC: Towards effective student authentication. In *5-th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2013)*, pages 289 – 292, Xian, China. IEEE Computer Society. 10.1109/INCoS.2013.52.

Miguel, J., Caballé, S., and Xhafa, F. (2015d). Methods and issues in visualization of trustworthy data from eLearning systems. In *Fifth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional approaches (ALICE 2015)*, Taipei, Taiwan. IEEE Computer Society. submitted.

Miguel, J., Caballé, S., and Xhafa, F. (2015e). *Intelligent Data Analysis for e-Learning: Trustworthiness for Enhancing Security in on-line Learning Teams and Networks*. Intelligent Data-Centric Systems. Elsevier, Amsterdam, Netherlands. Book proposal submitted.

Miguel, J., Caballé, S., Xhafa, F., and Prieto, J. (2014a). Security in online assessments: Towards an effective trustworthiness approach to support e-learning teams. In *28th International Conference on Advanced Information Networking and Applications (AINA 2014)*, pages 123–130, Victoria, Canada. IEEE Computer Society. 10.1109/AINA.2014.106. **Best Paper Award of AINA 2014**.

Miguel, J., Caballé, S., Xhafa, F., and Prieto, J. (2015a). Security in online web learning assessment. providing an effective trustworthiness approach to support e-learning teams. *World Wide Web Journal (WWWJ)*. Springer. doi:10.1007/s11280-014-0320-2. IF: 1.623, Q1: 20/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE).

Miguel, J., Caballé, S., Xhafa, F., and Prieto, J. (2015b). A massive data processing approach for effective trustworthiness in online learning groups. *Concurrency and Computation: Practice and Experience (CCPE)*, 27(8):1988–2003. Wiley Online Library. doi:10.1002/cpe.3396. IF: 0.784, Q2: 50/102, Category: COMPUTER SCIENCE, THEORY & METHODS (JCR-2013 SE).

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014b). A collective intelligence approach for building student's trustworthiness profile in online learning. In *2014 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet*

Computing (3PGCIC-2014), pages 46–53, Guangzhou, P.R. China. IEEE Computer Society. 10.1109/3PGCIC.2014.132.

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014c). A methodological approach to modelling trustworthiness in online collaborative learning. In *Fourth International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2014)*, pages 451–456, Salerno, Italy. IEEE Computer Society. 10.1109/INCoS.2014.18.

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014d). Predicting trustworthiness behavior to enhance security in on-line assessment. In *2014 6th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2014)*, pages 342–349, Salerno, Italy. IEEE Computer Society. 10.1109/INCoS.2014.19.

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2014e). Towards a normalized trustworthiness approach to enhance security in on-line assessment. In *Eighth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2014)*, pages 147–154, Birmingham, UK. IEEE Computer Society. 10.1109/CISIS.2014.22.

Miguel, J., Caballé, S., Xhafa, F., Prieto, J., and Barolli, L. (2015c). A methodological approach for trustworthiness assessment and prediction in mobile online collaborative learning. *Computer Standards & Interfaces (CSI)*. Springer. doi:10.1016/j.csi.2015.04.008. IF: 1.177, Q2: 38/105, Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING (JCR-2013 SE).

Miguel, J., Caballé, S., Xhafa, F., and Snasel, V. (2015). A data visualization approach for trustworthiness in social networks for on-line learning. In *29th IEEE International Conference on Advanced Information Networking and Applications (AINA-2015)*, pages 490–497, Gwangiu, South Korea. IEEE Computer Society. 10.1109/AINA.2015.226.

Moodle (2012). Moodle security announcements.

Mwakalinga, J., Kowalski, S., and Yngstrom, L. (2009). Secure e-learning using a holistic and immune security framework. In *Internet Technology and Secured Transactions, 2009. ICITST 2009. International Conference for*, pages 1 –6.

OrbiTeam (2015). BSCW groupware for efficient team collaboration and document management:.

Parker, D. (2002). Toward a new framework for information security. In *The Computer Security Handbook*. John Wiley & Sons, Inc., New York, 4th edition.

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., and Bragge, J. (2006). The design science research process: a model for producing and presenting information systems research. *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)*, pages 83–106.

Píriz, S., Gumbau, J. P., and Jiménez, T. (2013). Universitic 2013: situación actual de las TIC en el sistema universitario español. In *Conferencia de Rectores de las Universidades Españolas (CRUE)*.

Raina, K. (2003). *PKI security solutions for the Enterprise : solving HIPAA, E-Paper Act, and other compliance issues*. Wiley Pub., Indianapolis, Ind.

Raza, M., Hussain, F. K., and Hussain, O. K. (2012). Neural network-based approach for predicting trust values based on non-uniform input in mobile applications. *Comput. J.*, 55(3):347–378. doi:10.1093/comjnl/bxr104.

Schneier, B. (2003). *Beyond fear : thinking sensibly about security in an uncertain world*. Copernicus Books, New York.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., and Gleave, E. (2009). Analyzing (social media) networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies*, C&T '09, pages 255–264, New York, NY, USA. ACM. 10.1145/1556460.1556497.

Soboroff, I. and Nicholas, C. (2000). Collaborative filtering and the generalized vector space model (poster session). In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, pages 351–353, New York, NY, USA. ACM. 10.1145/345508.345646.

Song, W., Phoha, V., and Xu, X. (2004). An adaptive recommendation trust model in multiagent system. In *Intelligent Agent Technology, 2004. (IAT*

2004). Proceedings. IEEE/WIC/ACM International Conference on, pages 462–465. 10.1109/IAT.2004.1342996.

Strijbos, J.-W., Martens, R. L., Prins, F. J., and Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers Education*, 46(1):29 – 48. doi:10.1016/j.compedu.2005.04.002.

Trustwave (2014). 2014 trustwave global security report. Technical report, Trustwave.

Von Solms, B. S. H. (2010). A model for information security in e-learning management systems. Dubai, UAE.

Weippl, E. R. (2006). Security in e-learning. In Bidgoli, H., editor, *Handbook of information security Vol. 1, Key concepts, infrastructure, standards and protocols.*, volume 1, pages 279–293. Wiley, Hoboken, NJ.

West, R. (2008). The psychology of security. *Commun. ACM*, 51(4):34–40. doi:10.1145/1330311.1330320.

Zhai, Z. and Zhang, W. (2010). The estimation of trustworthy of grid services based on neural network. *JNW*, 5(10):1135–1142.