
© 2014 IEEE. Es permet l’ús personal d’aquest material. S’ha de demanar permís a l’IEEE per a qualsevol altre ús, incloent la reimpressió/reedició amb fins publicitaris o promocionals, la creació de noves obres col·lectives per a la revenda o redistribució en servidors o llistes o la reutilització de parts d’aquest treball amb drets d'autor en altres treballs.

(c) 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.
A Methodological Approach to Modelling Trustworthiness in Online Collaborative Learning

Jorge Miguel\textsuperscript{1}, Santi Caballé\textsuperscript{1}, Fatos Xhafa\textsuperscript{1}, Josep Prieto\textsuperscript{1}, Leonard Barolli\textsuperscript{2}

Department of Computer Science, Multimedia, and Telecommunication
Open University of Catalonia
Barcelona, Spain
\{jmmoreno, scaballe, fxhafa, jprieto\}@uoc.edu

\textsuperscript{2} Fukuoka Institute of Technology, Department of Information and Communication Engineering
Fukuoka, Japan
barolli@fit.ac.jp

\textbf{Abstract}— Trustworthiness and technological security solutions are closely related to online collaborative learning as they can be combined with the aim of reaching information security requirements for e-Learning participants and designers. In this paper, we justify the need of trustworthiness models as a functional requirement devoted to improve 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 online collaborative activities through trustworthiness assessment and prediction.

\textbf{Keywords}— information security; trustworthiness; assessment, prediction, on-line collaborative learning.

\section{Introduction}

Computer Supported Collaborative Learning (CSCL) has become one of the most influencing educational paradigms [1]. In this context, security is considered as a significant factor with the aim of ensuring information managed in CSCL [2]. However, there exist still relevant drawbacks that impede e-Learning providers to provide and reach security requirements defined by e-Learning stakeholders. Even most advanced security solutions have drawbacks that cannot be solved with the Information and Communication Technologies (ICT) alone [3]. Hence, we propose a hybrid model based on technological and functional models, namely, a trustworthiness approach devoted to improve security in CSCL by building a trustworthiness methodology to offer guidelines for designing and managing security in online collaborative activities through trustworthiness assessment and prediction. To this end, we first review, in section II, the main works in the literature on security in CSCL, how trustworthiness is related to security, and trustworthiness methodologies. In section III, 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 IV, we concrete the most significant aspects in terms of specific methods through their application in real online courses. Finally, conclusions and further work are presented.

\section{Background}

In this section, we review main works in the literature on security in collaborative learning, how trustworthiness is related to security and trustworthiness methodologies.

\subsection{Security in Online Collaborative Learning}

According to [1] Computer Supported Collaborative Learning (CSCL) has become one of the most influencing educational paradigms devoted to improve e-Learning. Some authors have argued that information security has to be considered with the aim of ensuring information managed in CSCL; in addition, several technological solutions have been proposed [2],[4]. 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 does not offer an overall security solution [5],[6]. One of the key strategies in information security is that security drawbacks cannot be solved with technology solutions alone [3]. Even most advances security ICT solutions have drawbacks that impede the development of complete ICT security frameworks.

\subsection{Trustworthiness in e-Learning and Information Security}

To overcome security deficiencies discussed above, we have researched into enhancing technological security models with functional approaches [7],[8],[9]. In [10], 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 [11], through the study of the most relevant existing trust models, trustworthiness modeling can be classified into trustworthiness assessment and prediction models (note that in the literature on trustworthiness modeling the terms determination and estimation are also used to refer assessment and prediction respectively). In this paper, we considered both purposes of trustworthiness. In addition, we also consider trustworthiness models, rules, factors and features that we have discussed in [7],[8],[9] with the aim to enhance security in e-Learning through trustworthiness methods.
C. 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 [12], propose a generic methodology, called Trustworthiness Measurement Methodology (TMM), which 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 business processes, but the key concept of this methodology is interaction between agents, that 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: (i) Determine the context of interaction between the trusting agent and the trusted entity; (ii) Determine the criteria involved in the interaction; (iii) Develop a criterion assessment policy for each criterion involved in the interaction; and (iv) Determine the trustworthiness value of the trusted entity in the given context and time slot corresponding to the time spot of interaction by making use of specific metrics. In [13], the authors presented the foundations of formal models for trust in global information security environments, with the aim of underpinning the use of trust-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, building on concepts from trust management and domain theory, feature at the same time a trust and an information partial order. The formal model is focused on three target aspects, namely, trust involves entities, has a degree, is based on observations and determines the interaction among entities. In addition to methodology and formal model approaches, in another work [14], it is presented a trust architecture by introducing a basic trust management model based on trustworthiness previous modeling work.

III. TRUSTWORTHINESS AND SECURITY METHODOLOGY APPROACH

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

A. Theoretical Analysis

In these sections, we present our methodological approach called Trustworthiness and Security Methodology in CSCL (TSM-CSCL). For the sake of simplicity, the acronym TSM is used. TSM is a theoretical approach devoted to offer a guideline for designing and managing security in 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 in TSM form a TSM design and deploy cycle. Each cycle corresponds to an interaction over the 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 have discussed in our previous research [7],[8],[9].

TSM aims to deliver solutions for e-Learning designers and 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: (i) define security properties and services required by e-Learning designers; (ii) build secure CSCL activities and to design them in terms of trustworthiness; (iii) manage trustworthiness in learning systems with the aim of modeling, predicting and processing trustworthiness levels; and (iv) 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: (i) Students, as the main actors in the collaborative learning process and as targets of the trustworthiness analysis; (ii) Designers, that represent the role in charge of all e-Learning analysis and design tasks; and (iii) Managers, that develop management processes, such as deployment, monitoring or control tasks.

B. Methodology Key Phases

As shown in Fig. 1, the TSM methodology is divided into three sequential phases: (i) Building Trustworthiness Components integrated into the design of secure collaborative learning activities; (ii) Trustworthiness Analysis and Data Processing based on trustworthiness modeling; and (iii) Trustworthiness Assessment and Prediction to detect security events and refine the design process. Although we have assessed 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.

Although we have defined a 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 enhancements which will be deployed in the next deployment (i.e. the next time that the students 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, this impediment can be overcome through design changes that are deployed in the next execution cycle.

The rest of this section presents details of the processes in each phase of the methodology.

C. 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, and in addition, it 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. 1. Due to the first goal of the methodology is to design the trustworthiness component; we divide this phase into the following analysis considerations: (i) collaborative learning activities generate a significant amount of interactions. Due students’ interactions are closely related to trustworthiness modeling, designers have to consider and analyze each interaction which may be related to trustworthiness; (ii) analyze 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 [8] which can be defined as those behaviors that reduce or build trustworthiness in a collaborative group and 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; (iii) 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 [5]. Through selecting and analyzing 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 have 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 have presented a case dealing with one collaborative activity only, which generate its peer-to-peer assessment component; but this case may be extended to a set of collaborative activities implemented in one or several peer-to-peer components 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 have suggested the need of modeling 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 modeling tools. To overcome this impediment, we have reviewed the Educational Modeling Language (EML) that, with the indications presented in [5], allows designers to tackle with modeling security, CSCL activities and interactions.

D. Trustworthiness Analysis and Data Processing

So far, the e-Learning designer has built the trustworthiness component and it has to 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, and their implementation may require additional technological solutions such as normalization processes presented in this section.

Trustworthiness modeling and normalization processes in TSM, as can be seen from the diagram in Fig. 3, are based on the key concepts presented in the rest of this subsection.
(further information and details of these concepts can be found in our previous research [7],[8],[9]).

We introduced the concept of Trustworthiness Indicator as a measure of trustworthiness factors. Trustworthiness factors have been presented as those behaviors that reduce or build trustworthiness in a collaborative activity and they have been 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 [8].

<table>
<thead>
<tr>
<th>2. Trustworthiness Analysis and Data Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployed Components</td>
</tr>
<tr>
<td>TW Data Sources</td>
</tr>
<tr>
<td>Trustworthiness Modeling</td>
</tr>
<tr>
<td>TW Indicators</td>
</tr>
<tr>
<td>Normalization</td>
</tr>
<tr>
<td>Data Processing</td>
</tr>
<tr>
<td>Preprocessing</td>
</tr>
<tr>
<td>Parallel Processing</td>
</tr>
<tr>
<td>Trustworthiness Processed Data</td>
</tr>
</tbody>
</table>

**Figure 3. Phase 2: Trustworthiness Analysis and Data Processing.**

Regarding Normalization Functions, there are several reasons that impede the management and processing of trustworthiness levels directly. Among them, we can highlight several factors, 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 [9].

Once trustworthiness modeling concepts have been defined, the task of data processing starts, and 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 is a prerequisite for trustworthiness data processing. In addition, with regarding to computational complexity, extracting and structuring trustworthiness data is a costly process and the amount of basic data tends to be very large [7]. Therefore, techniques to speed and scale up the structuring and processing of trustworthiness basic data are required (see [7] for a parallel implementation approach to be developed in the context of trustworthiness data processing).

**E. 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 B.

Trustworthiness assessment and prediction stem from the analysis of the time factor in trustworthiness. Fig. 4 shows how trustworthiness assessment and prediction begins with the conversion of processed data into trustworthiness sequences by considering 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.

**Figure 4. Phase 3: Trustworthiness Assessment and Prediction.**

Once trustworthiness sequences have been built, the e-Learning manager is able to set out predictions and assessment processes. As presented in [8], methods intended to predict and assess trustworthiness are available in the context of peer-to-peer assessment and 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 modeling 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 to 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 analyzed by the tutor before taking any decision.

Finally, trustworthiness decision information is available, and then, e-Learning managers can analyze valid and useful information devoted to report security events, improve the framework design, or manage security enhances.

IV. EVALUATION

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

A. Real Online Courses

We have carried out two studies based on real online courses at the Open University of Catalonia\(^1\), 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, the 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 that 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 of the group was evaluated by the coordinator. General e-Learning activities were supported by a standard LMS, which offered both rating systems and general learning management indicators. Given the low number of students, we could study the data in much more detail as manually allowing us to experiment with several design alternatives and adapting the model to the design cycles proposed in TSM (see Section 3).

The second study extended the scope of the first one to a more standard scenario in which we could not manage so much flexibility and manual processes. Students’ evaluation was based on a manual continuous evaluation model by using several manual evaluation instruments. Manual evaluation was complemented with automatic methods, which represented up to 20 percent of the total student’s grade. Therefore, we implemented a hybrid evaluation method by combining manual and automatic evaluation methods, and the model allows us to compare results in both cases. 59 students performed a subjective peer-to-peer evaluation [15], that is, each student was able to evaluate the rest of class peers in terms of knowledge acquired and participation in the class assignments. The evaluation was performed for each course stages, which were taken as time references in trustworthiness history sequences.

B. 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 is received. The second activity of the component was a students’ forum (F), which created a collaborative framework devoted to enhance 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: (i) Surveys; (ii) Ratings; (iii) Students reports; and (iv) LMS indicators. To sum up, each instrument is integrated into the collaborative activity and it manages its own data formats.

C. Analysis and Data Processing with TSM

We have analyzed research instruments data formats in terms of data sources in TSM and, for each case; we have 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 [9]:

\[
N(\text{tw}_{r,q,m,s}) = \sum_{i=1}^{n} \frac{\text{tw}_{r,q,m,i}}{n-1}, \quad i \neq s
\]

where \(\text{tw}_{r,q,m,s}\) is the responses (R) indicator, \(s\) is the target student (i.e. the student evaluated); \(n\) 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 \(L_{tw_i}^n\), we have managed several indicators composition. The most suitable level in both courses is based on a weight model:

\[
L_{tw_i}^n = \sum_{i=1}^{n} \frac{(\text{tw}_i + \text{w}_i)}{n}, \quad i \in I, w = (w_1, \ldots, w_i, \ldots w_n), \sum_{i=1}^{n} w_i = 1
\]

where \(n\) is the total number of trustworthiness indicators and \(w_i\) is the weight for the normalized indicator \(\text{tw}_i\).

Regarding data processing, we have experimented with sequential and parallel implementations [7]. Sequential approaches have been feasible to manage data sources from several activities, such as responses in a survey or number of

\(^1\) http://www.uoc.edu
posts in a forum. However, due to performance issues, we have not been able to process log data from our LMS with a sequential approach. For this reason, we endowed our trustworthiness framework with parallel processing facilities. To this end, we designed a MapReduce algorithm [7] implemented in an Apache Hadoop\textsuperscript{2} and deployed in the RDLab\textsuperscript{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 [7] for the whole results).

D. Assessment, Prediction and Evaluation with TSM

Peer-to-peer components have been 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 evaluation 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 have experimented with a hybrid validation approach by combining manual, automatic, external, and internal validation methods. As an example of this model, we have analyzed similarity between manual evaluation results and automatic trustworthiness levels. The method to tackle similarity proposed is based on Pearson correlation [16].

Finally, we have considered two different methods to deal with prediction. The first approach may be based on neural networks [11] and the second one on collaborative filtering. 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.

V. CONCLUSIONS AND FURTHER WORK

In this paper, we first motivated the need to improve information security in online 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 has been evaluated by presenting specific methods and techniques applied to real online courses. As ongoing work, we plan to continue the methodology testing and evaluation processing 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 and indicators.

ACKNOWLEDGMENT

This work was partly funded by the Spanish Government through projects TIN2011-27076-C03-02 “CO-PRIVACY” and CONSOLIDER INGENIO 2010 CSD2007-0 004 “ARES”; and the project TIN2013-46181-C2-1-R Computational Models and Methods for Massive Structured Data (COMMAS).

REFERENCES


\textsuperscript{2} http://hadoop.apache.org
\textsuperscript{3} http://rdlab.lsi.upc.edu