Miguel, J. [et al.] (2014) Security in online learning assessment towards an effective trustworthiness approach to support E-learning teams. *2014 IEEE 28th International Conference on Advanced Information Networking and Applications: IEEE AINA 2014, 13-16 May 2014, University of Victoria, Victoria, Canada: proceedings*. IEEE. Pp. 123-130. Doi: http://dx.doi.org/10.1109/AINA.2014.106.

Security in Online Learning Assessment

Towards an Effective Trustworthiness Approach to Support e-Learning Teams

Jorge Miguel[1], Santi Caballé[1], Fatos Xhafa[2], Josep Prieto[1]
[1]Department of Computer Science, Multimedia, and Telecommunication
Open University of Catalonia, Barcelona, Spain
{jmmoneo, scaballe, jprieto}@uoc.edu
[2]Department of Languages and Informatic Systems
Technical University of Catalonia, Barcelona, Spain
fatos@lsi.upc.edu

Abstract—**This paper proposes a trustworthiness model for the design of secure learning assessment in on-line 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 trustworthiness model is designed in order to conduct the guidelines of a holistic security model for on-line collaborative learning through effective trustworthiness approaches. In addition, since users' trustworthiness analysis involves large amounts of ill-structured data, a parallel processing paradigm is proposed to build relevant information modeling trustworthiness levels for e-Learning.**

Keywords- ***trustworthiness; e-assessment; information security; collaborative learning; parallel processing***

I. INTRODUCTION

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

Although online 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 complete 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, new models are needed and, in this paper, we propose a trustworthiness approach based on hybrid evaluation which can complete these lacks and support e-assessments requirements.

Finally, in order to provide effective and just-in-time trustworthiness information from LMS, it is necessary a constant 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. Due to the large or very large size of log files generated daily, the massive data processing is a foremost step in extracting useful information and may require computational capacity beyond that of a single computer (i.e. sequential data processing). To this end, this work also studies, the viability of a parallel approach for processing large log data files of a real LMS using distributed infrastructures.

The paper is organized as follows. Section II shows the background and contextual about security in e-Learning. Section III 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 IV, we extend our security model with the trustworthiness dimension. Once studied trustworthiness factors and rules, in section V we describe an innovate model based on trustworthiness applied to e-assessments. Since users' trustworthiness analysis involves large amount of data, parallel processing paradigms are proposed in section VI to build relevant information modeling trustworthiness levels. Finally, Section VI concludes the paper highlighting the main ideas discussed and outlining ongoing and future work.

[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 online 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

II. 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 [1] are focused on confidentiality and theses privacy approaches can be found in [2]. Despite the privacy relevance in secure e-Learning, it is important not to forget that information security is not only privacy services, and in further works [3], [4] security in e-Learning has been treated following more complex analysis and design models.

In [4] the author argue 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 learning management systems as a general key in e-Learning design and management. Furthermore, in [3] 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, 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 [2], [5]–[7].

Although the authors who have been mentioned so far, discuss security design in e-Learning from a theoretical point of view, other ones have argued that we actually need to understand attacks in order to discover security design factors we need to put into place and it is also needed in order to figure out how security services must be designed [8]. Researchers have already conducted many efforts proposing taxonomies of security attacks. In [9], through analyzing existing research in attack classification, a new attack taxonomy is constructed by classifying attacks into dimensions, this paper is mentioned because, besides the new taxonomy proposed by the authors, this work offers 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 [2], we have proposed an alternative approach which associate attacks to security design factors.

Furthermore, we still need 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 because if the reality of attacks is not significant today, our research would not be relevant. In this sense it can be found several reports which justify the relevance of security attacks during the last two years, in concrete terms, the study presented in [10] uncovered that security attacks are a reality for most organizations: 81% of respondents' organizations experienced a security event (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 [2], [5]–[7] 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. Although these approaches are suitable and are the source of our current investigation, with the purpose of enhance security in e-Learning, in the next section we conduct our investigation on one of most relevant topics which arise when information security and e-Learning are related and analyzed together.

III. Secure e-assessments

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, examining and selecting most relevant ones, then an assessments classification is depicted in order to analyze how e-assessments types and factors are related to previously selected security properties and, eventually, we propose a security model which extends technological security techniques adding functional requirements to secure e-assessments.

A. *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 these 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, description of activity, etc.) must not be changed, destroyed, or lost in an unauthorized or accidental manner. The result delivered by the student must carry out integrity property too
- Identification and authentication. While performing the evaluation task, the fact that students are who they say they are must be verifiable in a reliable way. In addition, both students' outcomes and evaluation results must actually correspond to 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. Selected properties are identity and integrity. Integrity must be considered both as authorship as well as integrity data.

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

Therefore, we will be able to trust in an e-assessment process when identity and integrity properties are accomplished. In the context of e-assessments, with regarding to identity, students are who they say they are when, in an assessment process; 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.). And dealing with integrity and authorship, we trust in 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 identity. Therefore, services applied to e-assessment must be considered in both a static and a dynamic way.

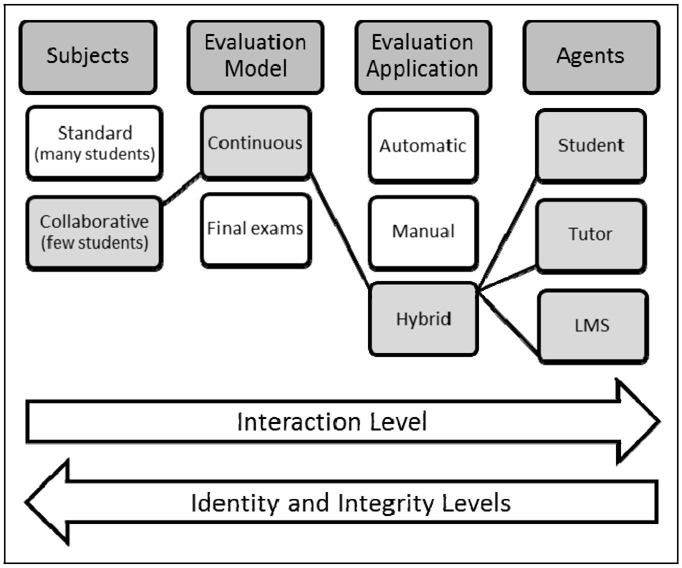

Figure 1. Evaluation types

B. Assessments Classification

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. Fig. 1 shows factors and evaluation types.

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' identity 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' identity 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 identity 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 Fig. 1 it has been marked the marked those elements which are involved in the model proposed. In the following sections it is presented how the secure e-assessment model is defined.

C. Beyond Technological Approaches

According to [11] problems encountered in ensuring that 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) tool are not enough to 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 [12] 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 e-exam statement at the same time; the exam is based on a statement that 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 his work into a digital document with his own resources. When 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' identity and digital signature for outcomes integrity and authorship. Therefore, the process described above is adapted in 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 statement of the e-exam and begins his work.
- The student checks the built-in digital signature statement in order to validate the integrity of this element.
- When the student finishes his work in the outcomes document, the student performs the operation of digital signature (into the digital document and using his 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 identity 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 that we expect or the student has send this resource to another one. Although we can add additional technological measures such as certificate storage devices, either cryptographic or digital file card protected by hardware or symmetric cryptography. There are ways to export these keys or allow remote access to devices which manage them, allowing these operations and, therefore, we can conclude that the student may share their resources identification and signature.

IV. TRUSTWORTHINESS

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 first extend our security model with a trustworthiness dimension.

A. Trustworthiness and Security Related Work

In [13] 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. On the other hand, as it is stated by the author, security is also a feeling, based not on probabilities and mathematical calculations, but on your psychological reactions to both risks and countermeasures. Since this model consider two dimensions in security and being aware that absolute security does not existed (it has been justified in the section "Beyond Technological Approaches") it can be stated that any gain in security always involves trade-off, even as it is concluded by the author, all security is a trade-off. 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. This trade-off is proposed because, as it is concluded by the author, we need both be and feel secure.

As it has been presented, 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. Therefore, we are going to continue our related work study taking general trustworthiness references.

B. Trustworthiness Factors

Beyond the overview of security and trustworthiness presented, we need to review how trustworthiness can be measure and which are the factors involved in its quantitative study. In [14] it is proposed a data provenance trust model which takes into account factors that may affect the trustworthiness and, based on these factors, assigns trust scores to both data and data providers.

In our context, students and students' resources (such a document, a post in a forum, etc.) can be modeled following this approach. Moreover, factors that may affect trustworthiness when students are developing collaborative learning activities must be discovered. To this end in [15], the author design a survey to explore interpersonal trust in work groups identifying trust-building behaviors ranked in order of importance. We use these behaviors 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 the following table:

	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 for .
4	S jumps to conclusions without checking the facts first.
5	S makes excuses or blames others when things do not work out.

a. Trustworthiness Factors

C. Trustworthiness Rules and Characteristics

Trustworthiness levels may be represented as a combination of trustworthiness factors which has been presented. Moreover, according to [16] there are different aspects of considering on trust, 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 trustworthy factors and rules which have been presented in this section. Furthermore, we will define two additional concepts (trustworthiness levels and indicators) which will be presented in the following sections.

D. Evidences and signs

Trustworthiness factors are defined from the perspective of students' behaviors 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 it allows us to state that a certain security property has been venerated. 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 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 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 analyzing 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.

V. A TRUSTWORTHINESS MODEL

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

A. 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 both to collecting 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.

B. 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, reports, etc.). 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 regarding to trustworthiness rules, this indicator may be compared to the group, over the time or considering the context. Trustworthiness indicators can be represented following these expressions:

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

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 above 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 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 [17] 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:

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

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.

C. Statistical Analysis

Following the trustworthiness model presented we need to inquire 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 with regarding to rates of similarity, correlation or dependence between two variables [18]. Even though the scope 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 (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:

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

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, anomalous behavior, 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 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 de 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:

$$r_a = (r_{a,1}, \dots, r_{a,i}, \dots, r_{a,n-1}), i \in I, i \neq a$$

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 pervious measures, adding time factor variable:

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

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 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:

$$r_{a,t} = (r_{a,1}, \dots, r_{a,i}, \dots, r_{a,n-1}), i \in I, i \neq a$$

Trustworthiness indicators which have already presented in this section are summarized in the following table:

Basic Indicators	**Trustworthiness Statistical Analysis**		
	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

b. Trustworthiness Basic Indicators

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

$$r_{a,b=cv_t}$$

where the second indicator (b) is exchanged by the value in continuous evaluation. According to this indicator, we will be able to analyze 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, it has been mentioned that, in the case of questionnaires, some questions which evaluate the same trustworthiness factor, are proposed in two different ways: individual and group evaluation, that is, students are asked about some factors related to their every member in the 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:

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

Indicators which have been presented in this section will be studied in the analysis stage of the model. Although they are proposed as suitable options, the model will be 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.

VI. Parallel Processing Approaches

According to [19] extracting and structuring LMS 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 collaborative learning. With regarding to computational complexity, extracting and structuring LMS data is a costly process and the amount of log data tends to be very large. Therefore, techniques to speed and scale up the structuring and processing of log data are required dealing with log data. In this section we present a guideline of parallel implementations which may be developed in the context of data logs of LMS.

A. *Parallelizing log files processes*

In [20] it is studied the viability of processing large log data files of a real virtual campus (UOC Virtual Campus) using different distributed infrastructures to examine the time performance of massive processing of log files. These models were implemented following the master-slave paradigm and evaluated using Cluster Computing and PlanetLab platforms.

Taking below approaches as starting point, we extend their goals in two different ways which are presented in next sections: parallelizing the normalization of several LMS logs files and using MapReduce paradigm.

B. *Parallelizing and normalization*

In [19] the task of structuring event log data can be defined as the processes which give 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 an each source. Each log file has its own format, otherwise, we cannot consider neither unified nor normalized those logs generated by the same Web Server; hence stage 1 is needed in order to normalize these sources following a unified format. We propose the following tuple:

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

which represents an user u performing 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 log files, the log data analysis of the information captured may be performed in two either in sequential or parallel. The most common approaches suggest that analyzing process should be developed in sequential processing using specific programs for statistical analysis. Moreover, statistical analysis computations are not usually suitable in parallel processing. Otherwise, we would consider the parallel processing approaches if the amount of data was large enough and the computation cost (calculating the statistical result) was low and elemental (e.g. counting number of items).

C. *MapReduce Paradigm*

The parallel implementation in the distributed infrastructures that we propose in this section follows the MapReduce paradigm. We introduce our MapReduce model parallelizing the normalization of different LMS log files.

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, log per day, etc.). The input may be represented as:

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

where L is the set of LMS, and I is the set of log files in a 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 Map phase and which ones are processed in Reduce, selecting the input and output keys for each phase. We have to be aware that 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 processes following the normalization process which has been presented and this output will be the input for the reduce function. At this point, we can decide among several alternatives dealing with reduce function. If we only want to store normalized data, the reduce task do not perform addition work, it only store the output of map function in the distributed file system. In addition, reduce function may be used to compute a relevant component in as it has been presented in the previous section. In this case, one of the keys is the student and reduce function calculate the result of the parameter selected (e.g. number of documents created by the student, total session time, sum of ratings, etc.).

D. *Hadoop*

With regarding to parallel platform supporting MapReduce paradigm, the abstract model proposed in above section will be implemented in Apache Hadoop[3]. In [21] it is presented the MapReduce 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.

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

In [21] a Hadoop MapReduce job is defined as a unit of work that the client wants to be performed: it consists of the input data, the MapReduce program, and configuration information. Hadoop runs the job by dividing it into tasks, of which there are two types: map tasks and reduce tasks. There are two types of nodes: a job tracker, which coordinates the paralleling process; and several workers which 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 governed by the size of the input.

The implementation of map and reduce function is based on these previous works [19], [20] which deal with different LMS log formats. Once the logs are computed by the event extractor functions, the output is normalized following the model presented. As a development environment we use Cloudera QuickStart VM[4]. This virtual machine contains a standalone Apache Hadoop framework with everything we need to test our model.

VII. Conclusions and Further Work

In this paper we have presented an innovative approach for modeling 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. This approach is based on trustworthiness factors, indicators and levels which allow us to discover how trustworthiness evolves into the learning system. We have proposed several research instruments for collecting students' data and since extracting and structuring LMS data is a costly process, a parallel processing approach has been proposed.

Ongoing work is implementing this abstract model in a real context. Therefore, we would like to select a real subject, among courses at UOC, and design an experimental pilot based on this subject. This pilot will extend the abstract model in a deployment framework offering a case study devoted to explore and enhance our trustworthiness model. Furthermore, in our future work, we will propose a benchmark for the each LMS used in the pilot, considering performance factors in those real paralleling platforms which have been proposed in this paper.

References

[1] S. K. Ferencz and C. W. Goldsmith, «Privacy issues in a virtual learning environment», en *Cause/Effect, A practitioner's journal about managing and using information resources on college and university campuses*, vol. 21, Educause, 1998, pp. 5-11.

[2] J. Miguel, S. Caballé, and J. Prieto, «Information Security in Support for Mobile Collaborative Learning», presentado en The 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS-2013), Taichung, Taiwan, 2013.

[3] C. J. Eibl, «Discussion of Information Security in E-Learning», Universität Siegen, Siegen, Germany, 2010.

[4] E. R. Weippl, «Security in E-Learning», en *Handbook of information security Vol. 1, Key concepts, infrastructure, standards and protocols.*, vol. 1, 3 vols., Hoboken, NJ: John Wiley & Sons, Inc., 2006, pp. 279-294.

[5] J. Miguel, S. Caballé, and J. Prieto, «Providing Security to Computer-Supported Collaborative Learning Systems: An Overview», presentado en Fourth IEEE International Conference on Intelligent Networking and Collaborative Systems (INCOS 2012), Bucharest, Romania, 2012, pp. 97-104.

[6] J. Miguel, S. Caballé, and J. Prieto, «Security in Learning Management Systems: Designing collaborative learning activities in secure information systems», *eLearning Papers. European Comission: elearningeuropa.info*, 2012.

[7] J. Miguel, S. Caballé, and J. Prieto, «Providing Information Security to MOOC: Towards effective student authentication», presentado en 5-th International Conference on Intelligent Networking and Collaborative Systems INCoS-2013, Xian, China, 2013.

[8] J. D. Demott, A. Sotirov, and J. Long, *Gray Hat Hacking, Third Edition Reviews*, 3.ª ed. New York: McGraw-Hill Companies, 2011.

[9] Z. Wu, Y. Ou, and Y. Liu, «A Taxonomy of Network and Computer Attacks Based on Responses», en *Information Technology, Computer Engineering and Management Sciences (ICM), 2011 International Conference on*, 2011, vol. 1, pp. 26 -29.

[10] CSO Magazine, US Secret Service, Software Engineering Insistute CERT Program at Carnegie Mellon University, and Deloitte, «2011 Cybersecurity Watch Survey», CSO Magazine, 2011.

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

[12] Y. Levy and M. Ramim, «A Theoretical Approach For Biometrics Authentication of E-Exams», presentado en Chais Conference on Instructional Technologies Research, The Open University of Israel, Raanana, Israel, 2006.

[13] B. Schneier, «The psychology of security», en *Proceedings of the Cryptology in Africa 1st international conference on Progress in cryptology*, Berlin, Heidelberg, 2008, pp. 50-79.

[14] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, «An Approach to Evaluate Data Trustworthiness Based on Data Provenance», en *Secure Data Management*, vol. 5159, W. Jonker and M. Petković, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 82-98.

[15] P. Bernthal, «A survey of trust in the workplace», HR Benchmark Group, Pittsburg, PA, Executive Summary, 1997.

[16] Y. Liu and Y. Wu, «A Survey on Trust and Trustworthy E-learning System», 2010, pp. 118-122.

[17] I. Ray and S. Chakraborty, «A Vector Model of Trust for Developing Trustworthy Systems», en *Computer Security – ESORICS 2004*, vol. 3193, P. Samarati, P. Ryan, D. Gollmann, and R. Molva, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 260-275.

[18] B. Mobasher, R. Burke, R. Bhaumik, and C. Williams, «Toward Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness», *ACM Trans. Internet Technol.*, vol. 7, n.º 4, oct. 2007.

[19] F. Xhafa, C. Paniagua, L. Barolli, and S. Caballé, «A Parallel Grid-Based Implementation for Real-Time Processing of Event Log Data of Collaborative Applications», *Int. J. Web Grid Serv.*, vol. 6, n.º 2, pp. 124–140, jun. 2010.

[20] S. Caballé and F. Xhafa, «Distributed-based massive processing of activity logs for efficient user modeling in a Virtual Campus», *Cluster Computing*, abr. 2013.

[21] T. White, *Hadoop: the definitive guide*, Third edition. Beijing: O'Reilly, 2012.

[4] http://www.cloudera.com