Development and Application of a Proxy Server for Transparently, Digitally Signing E-Learning Content

Christian J. Eibl¹, SH Basie von Solms², and Sigrid Schubert¹

¹ Didactics of Informatics and E-Learning, University of Siegen, Germany {eibl,schubert}@die.informatik.uni-siegen.de

² Academy for Information Technology, University of Johannesburg, South Africa basie@rau.ac.za

Abstract. Integrity as minimal requirement for successful protection of the learning process is neglected by most learning management systems. To implement integrity protection independent of existing e-learning systems we present the concept of a proxy server that digitally signs outgoing messages to the learning management systems and verifies the signature of incoming messages accordingly. We illustrate the architectural design and give specification details. Realization deliberations are outlined with respect to the hypertext transfer protocol. Finally, we discuss the approaches on the case study of the open-source learning management system MOODLE.

1 Motivation

E-learning as well as many other computer based services were concerned by the growing interconnection, i.e., security has become a major topic in this field. Even more, security as a quality factor is vital to the reputation and further existence of institutes, companies or similar being interconnected to thousands of users over the Internet. Security issues in learning management systems (LMS) that are already in extensive use will undoubtedly lead to discouragement of participating learners. That means that especially in commercial offerings or at distance universities that demand fees for their e-learning material the quality of such provided material and the e-learning system itself has to be best possible. Such quality must be due in courses one pays for.

Regarding different e-learning systems, it becomes obvious that security cannot be considered uniformly. We have to cope with different system types and their security requirements accordingly. Exemplified the following system types will be distinguished and described more in detail below:

- 1. simulation system,
- 2. virtual university,
- 3. assessment system.

In simulation systems, for example, learning material is read-only. There is no need for learners to be able to store data on the e-learning server. In virtual universities, on the other hand, all aspects of real universities get mapped to this piece of software, and therefore, it demands write-access on the e-learning server, e.g., for communication capabilities. As third type, an assessment system describes facilities that are used to assess learners in a computer-based way. Such a system requires a lot of security mechanisms to prevent learners from cheating or denying their participation. This type of system, of course, is the one with the highest security requirements in order to be incontestable.

For being able to ensure integrity, e.g., of learning content, as well as parts of non-repudiation, e.g., for proving the participation of some student, the digital signature gives an appropriate means – it even allows multiple signing of learning content to signalize consensus among teachers. But although digital signatures have a mainly accepted value for security, the application of such methods demands a lot of interaction with users. Especially when thinking ahead, i.e., if we would like to sign all messages sent to an LMS even those for chats and other communication capabilities, then digital signatures will become a weary load for users. Learners must not be disturbed more than necessary in their learning process by the system's security mechanisms in order to increase a positive learning outcome. Hence, we need an automatism that can cope with these aspects, i.e., the increased security by digital signatures and the automatic application of those methods.

This paper will present a proxy server running locally on every client machine that automatically appends digital signatures to messages sent to the LMS and, of course, evaluates incoming messages again according to their embedded signature.

2 Security in E-Learning

E-learning has become more and more focus of security investigations. An exhaustive security examination is given by the doctoral dissertation of Graf [6]. Graf considers different security issues concerning e-learning systems and focuses on secure assessment systems in WWW-based learning environments, i.e., using the World Wide Web (WWW). His approaches mostly depend on Java programs that establish communication between client and server instead of using the Hypertext Transfer Protocol (HTTP) which can be easier manipulated and eavesdropped. Weippl [14] as well as Graf provide an overview over the area of information security in e-learning, but Weippl also considers risk analysis approaches. He regards the security requirements from different points of view according to the corresponding role in the system. As roles he proposes authors, teachers, managers, and students. Teachers and authors, although in other literature often combined into the same person, are distinguished in that way, that authors are the ones who write the learning content, and therefore, are concerned with security of this material. Teachers, on the other hand, take place

while the system is already running. They coach students and manage different tasks like supporting students in their courses, administering resources, and doing exams. Managers have nothing (directly) to do with learning content and the learning process, but they are responsible for the superior administration and the security of the whole system. The authors' concern that "readers must be able to rely on the correctness of the content" [14, p. 15] will be addressed in this article.

Considering the different system types introduced above, it becomes obvious that security requirements differ not only out of the point of view of different roles, but also that different systems have different minimal security requirements. The maximal desired security for all systems, of course, will be an invincible and everlasting system, that cannot be attacked even with infinite computational power and time, i.e., it is perfectly secure. But since perfect security is not achievable in practice, a maximal security describing a stage of security that needs so much effort that the expense of attacking exceeds the worth of the gained success by far is considered as sufficient. For our further discussion the aspect of minimal security will be focused.

As system with minimal security requirements simulation systems, for example, demand integrity of learning content, which can be supported by corresponding access control mechanisms for the server where data is stored on. Since no write-access has to be given to some user, the authentication process can be optional. As system with medium security requirements virtual universities demand secure authentication and availability for some high percentage rate of time. Non-repudiation is desirable for communication capabilities, but not necessary. Confidentiality is necessary for personal data, supported by access control mechanisms, and optional for learning content. The confidentiality of learning content depends on the financial system of the study course. If it should be available to paying students only, then confidentiality is necessary, otherwise the contents may be accessible to guests, too. In assessment systems the highest security requirements must be demanded. Here minimal security requirements tend to be equal to the maximal ones as described before. No student may be able to cheat to his advantage or to the advantage of someone else. In addition, even teachers have to be prevented from cheating for some student respectively from changing some submitted answers unintentionally. The non-repudiation security service is of high value in this case, too, since no participating student may be in the position to deny the participation, because for example he was not well prepared. Considering the uploading of assignment solutions, integrity and non-repudiation become most important if those assignments have high value for marking students. No adversary may be able to alter other solutions unrecognized and no student should be able to deny such an upload.

Geuer [5] presented a signature system for WWW applications that signs the whole data part of the HTTP stream. He proposes an extension to the HTTP protocol that requires the adjustment of web server as well as web client to support this extension. As alternative to the web client adjustments the use of a proxy server was proposed. However, since in e-learning systems not only

the integrity of data while it gets transmitted is relevant but also the integrity of every single message stored on the server, this approach is insufficient for e-learning purposes.

But how can managers be sure that the e-learning system itself was created with a secure concept in mind? For such evaluations there exist several catalogues (cf. [2, 3, 7, 12]) that give a hint how secure concepts can look like, and therefore, users can check the system against such catalogues of criteria. When considering security concepts of existing and commonly used LMS, like for example the open-source LMS MOODLE [10] (cf. Section 6), then it becomes obvious that especially integrity and non-repudiation are not protected at all. Most security aspects depend primarily on the service of authentication. If this authentication is secure enough, then no unauthorized learner or teacher can manipulate data or even read such data. From the point of view of administrators, resp. managers, the security of such authentication mechanisms should be best possible. The problem within this case is that e-learning should be possible "anytime and anywhere you like" especially from the learners' point of view. When considering strong authentication mechanisms like sophisticated biometric systems there appear contradictory goals between the portability resp. flexibility of the system as demanded by learners and security as desired by administrators. Common LMS rely primarily on password authentication mechanisms which are, unfortunately, mostly implemented without sufficient support in secure password generation and regimentation. In MOODLE, for example, there exists neither a simplicity analysis of passwords nor checking mechanisms against dictionaries. Consequently, a lot of inexperienced students will most probably choose very simple passwords without being warned by the system.

Since the authentication process as "first line of defence" is conceptually hard to manage as sketched above because of the contradictory sight of learners and administrators, we propose to implement a "second line of defence". For protecting integrity and non-repudiation aspects digital signatures can be used (cf. [11]). With respect to learners' learning process an automatism for the signing process is needed which seem to be an optimal task for a proxy servers, since they give a simple framework for altering passing packets on the network.

In the following, we regard only the case of universities using e-learning capabilities. The approaches can be adopted to, e.g., schools or companies analogously.

3 Demands and Possibilities of Signing Proxy Servers

3.1 University wide PKI

First, there is the question of how to equip every user of the e-learning system with a public-private-key pair in order to digitally sign messages. Since public keys should be verified to belong to the pretended owner in order to satisfy authenticity, we can sign such public keys after checking the identity of the owner. In universities this process can, for example, be implemented as university wide public key infrastructure (PKI) [15]. In the following we will sketch very shortly a possible PKI implementation at a university without the demand of completeness.

Since at universities every student has to get registered at the central university administration, such an administration could take the role of the registration authority (RA) in the PKI. That means, after students have created their own pair of keys, the RA verifies the identity of every student registering in person and enables the creation of a certificate by the certification authority (CA). As CA we suggest the data processing centre or an executive of the university. Hence, every student gets a certificate that can be used for all university purposes.

For the sake of equality we have to cope with the situation that some students might not possess a computer on their own, and therefore, they need a terminal where they can generate their keys instead. Since every student needs some kind of student ID card, this can be, for example, combined with the idea of equipping every student with a smart card that contains the pair of keys on a chip and personal data of the student printed on the surface. Alternatively USB sticks could be offered by the administration for storing keys on it. This smart card or stick could then be used at all workstations in students' computer pools on the campus. The generation at terminals in the university administration would have the advantage that in case of problems competent personal is present to support. This will prevent discouragement in an early stage.

Note that there are a lot more advantages than only using digital signatures in the e-learning system, e.g., online registration for exams, which could motivate such an infrastructure additionally. However, in this work e-learning systems will be focused.

3.2 Proxy Server Architecture

The second problem is, how to establish automated security mechanisms on the client machine, i.e., how to secure the communication without disturbing the learning process. For this the use of a proxy server is sensible.

Considering the kind of messages sent to an e-learning system, i.e., learning content created by teachers or communication messages for forums and chats (by teachers and learners), it becomes obvious, that especially in the case of very short messages (like found in chats) the manually applied digital signature would mean a lot of inconveniences that enormously will interrupt the learning process and puts the sense of communication with other learners into question. The use of a proxy server as automated secretary that digitally signs all outgoing messages to the e-learning system is possible, since a proxy server runs logically on ISO/OSI (International Organization for Standardization / Open Systems Interconnection) layer 7, the application layer (cf. [9]). It is therefore able to examine and alter packets passing it, see Figure 1. The proxy server running on the client machine first has to gather access to the private key of the logged in user. This key can be, as described above, stored on a smart card, a USB stick or, if the pair of keys was created on the currently used computer, locally on

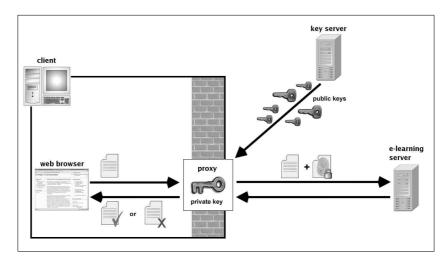


Fig. 1. Architecture with signature proxy; the LMS is running on e-learning system.

the harddisk drive. Since private keys should be protected by a pass phrase or other authentication methods, the proxy will have to demand the identification of the user. When the key could successfully be decrypted it needs to be stored in main memory to give the proxy the chance to use it several times without asking for authentication again at every access. Note that this implies some kind of security issue in favour of the learners' convenience. An attacker who can gain direct memory access would with some effort be able to extract that private key. Since direct memory access is on modern operating systems only possible by the super user account or by exploiting interface design problems (cf. [1]), this leads to administrative concerns and configuration issues that are outside the scope of this article. In the following we assume the systems to be well configured and administered to prevent the sketched attacks. To avoid this security issue even after the learner has quit the proxy server, we propose a secure deletion of the private key in main memory, i.e., in addition to freeing memory the location where the key has been stored must be overwritten (several times) by random data to scramble this information.

With holding the private key in memory the proxy can sign all messages to the e-learning system (cf. Figure 1). Of course, the proxy server needs some instructions on how to find the relevant messages and how to decide, whether the application of the signature is sensible and desired by the user. Such patterns for decision will be presented in Section 4.

For the other direction, i.e., incoming messages, the proxy server can take the role of the control instance for contained digital signatures. For evaluating embedded signatures it needs access to corresponding public keys. For this purpose a keyserver can be used that provides all known public keys of the university wide PKI.

4 Specifications

The implementation has to result in a small and fast program that is available for most platforms, i.e., a platform independent implementation is desired for not excluding some possible learners from using the more secure system. Besides this, the proxy server will have to cope with the following tasks presented more in detail below:

- 1. network adjustability,
- 2. LMS recognition,
- 3. LMS specific signing and verification,
- 4. communication with learner resp. teacher,
- 5. exclusion prevention.

1. Network adjustability: In association with the functionality of the network, the proxy server, of course, has to be able to cope with several simultaneous connections. That means, every connection established through the proxy server needs its own thread. Otherwise, the first connection would block the proxy server and no other connection could take place before the first one has been reset. For the sake of performance of the connection we propose the restricted storage of packets by forwarding data as fast as possible to the LMS. If storing the whole message before forwarding data to the meant target, i.e., elearning server, then this can be very time consuming when considering sending of large file attachments. Hence, data needs to be forwarded as fast as possible while the signature can be computed on a stored copy of the whole message and afterwards be included into the stream. For being able to use the proxy server in the wide variety of networks used out there, it has to be highly configurable. For example, if the network configuration is very restrictive by using special firewall rules or demanding other proxy servers in the network, then our proxy server must be able to run on different ports of the client as well as supporting chains of proxy servers to co-operate with demanded other ones.

2. LMS recognition: Since there are a lot of different LMS available on the market, the extendability of the proxy server is highly desired. The differences in LMS can be coarsely described by the tuple (serverURL,site). The first element describes the host URL (Uniform Resource Locator) for the LMS, while the second element describes the relative path on the target machine. For being able to recognize, whether a user tries to connect to a known LMS, the proxy server has to compare data like host URL and requested site from the received packet against patterns stored in its configuration files. If none of the patterns match to the currently requested page, the proxy server, of course, must not alter any detail of the request and has to forward packets without any delay, i.e., the network functionality must not be reduced in any way. Note that tasks 3 to 5 may only be applied in case of communicating with a known LMS, otherwise the communication would (at least slightly) be disturbed by

unnecessary computation of received packages without giving any advantage.

3. LMS specific signing and verification: Signing and verification of embedded signatures, of course, are two different subtasks. The first, i.e., signing messages sent to an LMS, needs information about variable names contained in packets sent to a specific site on the LMS. The proxy server needs such information in order to separate content to be signed within the observed HTTP stream from that data that can be dropped. In addition, the proxy server needs to know how to combine those variables to a new simple message that finally can be digitally signed. A list of variables appears sensible to mention the relevant variables in an ordered way.

The verification process, on the other hand, needs a lot more descriptive information. Consider the case, that variables sent to the server can only be requested again in form of a flat HTML (Hypertext Markup Language) page without forms or other rather dividing structures. The messages as presented by the LMS are no longer stored in variables, but are embedded in HTML page text. In the configuration files, data about the recognition of digital signatures embedded in those pages and the possible segmentation in single message blocks needs to be given. An example is described in case study MOODLE in Section 6.

4. Communication with learner resp. teacher: The communication capability with learner resp. teacher is a main task in order to be able to inform users adequately and to support their work with the signing system. Since the proxy server has to communicate in some scenarios like warning the user if an incoming message was altered fraudulently, this should preferably take place in the language of the learner for being as comfortable as possible. Consequently, the proxy server should be dynamically adjustable to the language of the host system.

For the way of communication there exists the possibility of building a graphical user interface or the approach of modifying the HTTP stream, i.e., the proxy server can add some message text or colour bar for example on top of the requested web page to signalize the security status of this site.

5. Exclusion prevention: Since security improvements by using the proxy server demand modification and evaluation of data packets passing this proxy server, it must be able to compute received data. This means, if client and server are trying to negotiate some encrypted connection or a compression (cf. Section 5) that the proxy server does not support, then this would lead to exclusion from communication. An example is given with the use of encrypted connections based on Secure Socket Layer (SSL) resp. Transport Layer Security (TLS). Since such end-to-end encryptions take place between client, i.e., web browser, and server, i.e., web server, the proxy server will only be confronted with an encrypted HTTP stream. For being able to cope with such a situation, the proxy server has to adopt the client part for this encryption. That means,

the proxy server is that party that negotiates encryption parameters with the server, and the web browser sends its data unencrypted to the (local) proxy server. Since this data will be sent only locally on the client machine, the lack of confidentiality by this now unencrypted connection is neglegible. Of course, the proxy server has to inform the user about server certificates used for the SSL/TLS encrypted connection. The proxy server must not be the deciding part, whether a certificate is trustworthy or not! But with the storage of most common certificates of certificate authorities, like web browsers do as well, it can support the user in his decision by verifying the server certificate in advance.

5 Realization

A main realization goal is the conformity to the protocol HTTP in version 1.1 (HTTP/1.1) as described in Request For Comments (RFC) 2616 [4]. There all details are described for the content of data sent from client to server and vice versa. First, the proxy server has to fulfil all requirements concerning the proxy functionality itself. The RFC 2616 describes different usage of address requests and their handling. This can be simply adopted from the standard and will therefore not further be discussed in this paper. Second, the proxy server has to be able to identify relevant parts of the HTTP stream and react appropriately. A main problem concerning the ability of recognizing relevant parts in the stream is that the stream itself is readable for the proxy server. HTTP/1.1 allows several ways of encoding and compression. Usually a web browser informs the web server about its supported compressions, the web server on the other hand compares this information to its own supported formats and decides accordingly. If the proxy server in the middle does not support the negotiated compression, then it will not be able to work with such data. Hence, the offerings of the web browser need to be considered as well and possibly be adjusted to the capabilities of the proxy server. Of course, such adjustments are only necessary and sensible if the connection does match a pattern as introduced in the former section.

Now that the proxy server can read all sent data in plain text it can search for the message to be digitally signed. For this the list of variables as described above can be used. With extracting all relevant variables and concatenating them to sign them all at once, the main task of the proxy can be fulfilled. Note that the list of variables must not be modified if it has been already used in the given way by any client, since otherwise the modification of the variable order or the number of variables would lead to another concatenated plain text, and therefore, will lead to another signature. Hence, with an altered list of variables old signatures would be falsified anytime they get checked with the new verification method description. Since HTTP supports different request methods affecting the composition of messages, we have to distinguish their handling. The most commonly used methods are POST and GET. Also relevant for the proxy server are messages encoded as multi-part messages. The proxy

server has to be able to extract the variable names and contents of all these kinds of messages. While in GET messages variables are coded into the URL, in POST messages the URL does not contain any variable, but the packet body does instead. In multi-part messages every variable gets its own field in the packet body, where these fields are separated by a boundary string introduced in the packet header. Hence, before processing the received data the type of message composition has to be determined.

To manage transparency for all users, it is sensible to store the signature in that way that it cannot be seen at first sight by a learner or teacher not using the described proxy server. For achieving this, the digital signature has to be stored in HTML tags that are not displayed. Since most LMS contain a content management system filtering user entries, there is no general solution how to hide those signatures. In the following section problems arising within MOODLE will be discussed.

6 Case Study MOODLE

Since MOODLE is now used for round about one year at the university of Siegen we could collect some experiences with its usage, implementation, and security concept. Consequently, when thinking about a prototype of the introduced proxy server we focus on the support of this LMS. It provides several learning activity capabilities like chat, forums, and polls that support learners in building social contacts within the LMS and provide platforms for discussion and exchanging messages. Of course, any of these learning activities gets controlled by a script page for reading and sending new messages. Since the functionality is still the same for, e.g., forums in different courses, the same script is used for all those courses with unique course identifier as parameter. Hence, only a finite number of sites must be observed, and therefore, the tuples representing search patterns can be easily created.

The verification of embedded signatures, e.g., in forum entries, can be handled as well. The forum gets presented as HTML page without form tags, but since forum entries have hierarchical structuring to separate main statements from their corresponding answers and comments, every entry is embedded in table data tags. Hence, starting from the digital signature the table data environment of the same hierarchical level can be extracted and be analyzed due to the entries considered in the signature. This list of entries is directly connected to the list of variables introduced in the signing process above.

When examining MOODLE the invisible embedding of digital signature turns out to be not as easy as thought at first. The client can only communicate over the interface MOODLE offers, i.e., sites for sending and storing messages. Every message sent to MOODLE gets controlled and if necessary modified. MOODLE does not allow to send HTML comments or to apply Cascading Style Sheets (CSS) instructions to some text in order to make it invisible. This, of course, is sensible from the point of view of MOODLE programmers, since the possibility of injecting arbitrary code, e.g., in CSS definitions, means an enormous security problem. However, a possible solution to hide the signature is to add it to an URL, e.g., in an image source tag, separated from the actual URL by a question mark. The signature will be in this case interpreted as parameter to the web site given by the URL. If the site referenced by the URL does not evaluate any parameter it will not be any problem. This script could, for example, deterministically response with a transparent image. Hence, nothing can be seen by users, but the signature is still hidden in the received HTML text.

7 Conclusion

We have introduced the idea of a proxy server for automated digital signature of messages to an LMS, since most LMS on the market lack the security of integrity and non-repudiation. With the use of the presented proxy server this lack of security can be fixed. When using the formerly mentioned catalogues of criteria for checking the security of such whole e-learning systems (including the client machine, too), then the proxy server increases the security by remarkable value. Integrity and non-repudiation were in the case without proxy server not protected at all, and with the use of it these services are protected by a respectable security method like digital signature.

The LMS MOODLE, unfortunately, was more problematic than thought at first. It is a content management system, and therefore, has to filter the transmitted HTML code, since otherwise executable code could easily be injected. But filtering even commentary code and CSS directives made it a lot more difficult to embed hidden signatures. HTTP needs a lot of attention due to its set of parameters. These parameters could easily exclude the proxy server from showing its strengths. Therefore, the proxy server must not allow parameters that could prevent it from analyzing the communication.

Since the proxy server is still under development, there are a lot of adjustments to be done. After a prototype working reliably with MOODLE can be released, we will extend the set of patterns to support other commonly used LMS like ILIAS [8] or WEBCT [13], too.

Limits of our approach get obvious when considering signing large multimedia files. In this case signing and verification processes will be very time consuming, and therefore, it will no longer be an almost invisible task on the client machine without disturbing the learning process. Another concern is given by the subjective reliability of signed content. Since signatures give the feeling that the signed content has been very carefully proved before signing it, this could be misleading for learners. The main goal of protecting integrity could be misunderstood by learners as guaranteeing correctness by exhaustively prove reading content prior to signing it. This needs to be well communicated, since minor changes will be necessary with high probability.

A main problem in the current approach is, that we consider the proxy server as optional extension to an existing learning environment. This program should

work nearly unnoticeable on the client machine, but as long as the usage is not obligatory, there will be most probably a lot of participants not using this system. Hence, the security improvements get unfortunately depleted. As solution an authentication method based on digital signatures is conceivable to force the usage. Since MOODLE, for example, does not provide a challenge-response authentication based on digital signatures, this still is part of further research.

References

- Becher M, Dornseif M, Klein CN (2005) FireWire all your memory are belong to us. CanSecWest Conference, Vancouver, Canada, http://cansecwest.com/ core05/2005-firewire-cansecwest.pdf [02-02-2007].
- Department of Defense (USA) (1985) Trusted Computer System Evaluation Criteria. Report DoD 5200.28-STD ("orange book").
- 3. Eibl CJ, von Solms BSH, Schubert S (2006) A Framework for Evaluating the Information Security of E-Learning Systems. Proc. of the 2nd International Conference on Informatics in Secondary Schools Evolution and Perspectives (ISSEP), Vilnius, Lithuania.
- Fielding R, Gettys J, Mogul J, Frystyk H, Masinter L, Leach P, Berners-Lee T (1999) Hypertext Transfer Protocol – HTTP/1.1, Request for Comments (RFC) 2616.
- 5. Geuer CH (1998) Entwurf, Realisierung und Verifikation eines Signatursystems für das Word-Wide-Web. Diploma thesis (German), University of Siegen.
- Graf F (2002) Lernspezifische Sicherheitsmechanismen in Lernumgebungen mit modularem Lernmaterial. Doctoral dissertation (German), University of Darmstadt.
- 7. Grobler CP (2003) A Model to assess the Information Security status of an organization with special reference to the Policy Dimension. Magister Thesis, Rand Afrikaans University.
- 8. Integriertes Lern-, Informations- und Arbeitskooperationssystem (ILIAS), online resource: http://www.ilias.de/ios/index-e.html [31-10-2006]
- ISO/IEC 7498-1 (1996) Information Technology Open Systems Interconnection – Basic Reference Model: The Basic Model. International Standard, corrected and reprinted version, Geneva, Switzerland.
- Modular Object-Oriented Dynamic Learning Environment (MOODLE), online resource: http://moodle.org [31-10-2006]
- 11. Schneier B (1996) Applied Cryptography: Protocols, Algorithms, and Source Code in C. Second Edition, Wiley.
- Swanson M (2001) Security Self-Assessment Guide for Information Technology Systems. National Institute of Standards and Technology (NIST), Special Publication 800-26.
- 13. WebCT, commercial LMS, online resource: http://www.webct.com [31-10-2006]
- 14. Weippl ER (2005) Security in E-Learning. Springer, New York.
- Xenitellis S (2000) The Open-source PKI Book, online, last modified: 23-07-2000, URL: http://ospkibook.sourceforge.net/ [17-03-2006]