DOI: 10.7250/eunis.2013.040

A PDF BASED DIGITAL SIGNED DOCUMENT FORMAT FOR INTEROPERABLE INSTITUTIONAL STRUCTURED DATA INTEGRITY

Luís A. Maia[1], Luís M. Valente[2], Manuel E. Correia[3], Lígia M. Ribeiro[4], Luís Antunes[5]

[1]CRACS & INESC-Porto LA, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 1021, 4169-007 Porto, Portugal , lmaia@dcc.fc.up.pt.
[2]Reitoria, Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal, lvalente@reit.up.pt.
[3]CRACS & INESC-Porto LA, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 1021, 4169-007 Porto, Portugal , mcc@dcc.fc.up.pt.
[4]Reitoria, Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal, lmr@reit.up.pt.
[5]CRACS & INESC-Porto LA, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 1021, 4169-007 Porto, Portugal , lfa@dcc.fc.up.pt.

Keywords
Student information systems, dematerialization, XAdES, PAdES, SignServer, non-repudiation, digital signature.

1. ABSTRACT

It is widely recognized that information systems constitute a key tool for the overall performance improvement of administrative tasks in academic institutions. However at their genesis lies a latent promise of a paper-less environment that stays most of the time unfulfilled due to the lack of appropriate digital document integrity and accountability mechanisms. Academic institutions are thus most of the time still relying on traditional security trust methods based on paper documents for signing and archiving critical documents. While this method delivers an inefficient, inconvenient and costly workflow, it is still a common method to provide some sort of workable verifiable integrity and accountability that is still considered to be appropriate for the digital data that is being managed by the institutional information systems. Paper based documents have been relying on physical signatures and stamping policies and the physical properties of paper and ink for their integrity and authenticity for a long time. However, the evaluation of a paper document signature or stamp is not a straight forward process. It requires the recipient to have a notarized copy of the signer's signature or stamp for comparison and requires handwritten signature evaluation training that is often beyond the scope of many office employee training. This can lead to situations where the level of credibility and integrity of paper based document is not adequate and makes the verification process entirely dependent on the administrative staff capacity of recognizing hand written signatures and puts too much trust on physical stamps, some of which are non-locally issued and thus very difficult to authenticate. In critical contexts this clearly is not enough to provide appropriate levels of non-repudiation and integrity for critical documents issued by institutions.

Digitally signed structured XML documents provide an interesting solution to this problem. Not only can the validation of the document be fully automatized and its integrity verifiable in real time by the information system, but it can also be implemented in such way that the information contained in such structured documents can be safely and more easily integrated into different information systems without human intervention, thus allowing for substantial cost reduction and leading to faster process work-flows with increased security and data quality.

In this paper we propose a PDF based document framework where any signed XML (PDF) document, produced by the institution can be at a later stage directly dematerialized and integrated into any compliant information system in a secure way while maintaining the information integrity and the ability to be self-verifiable. This framework involves the embedding of an encapsulated XAdES signed XML document with the information used on its production as an attachment to a PDF document with an institutional rendering visualization of the signed XML data. The attached XML document and the PDF are both time stamped by an external entity and signed by employees and the issuing institution.

DOI: 10.7250/eunis.2013.040

2. U.PORTO INFORMATION SYSTEM

The Students Information System currently being developed and in use by the University of Porto, called SIGARRA - Information System for the Aggregate Management of Resources and Academic registers[1], started its development in 1992 with the main objective to manage student records in the Faculty of Engineering offices evolving to a web oriented information system at a later stage.

With continuous development and growth in the number of functionalities, the information system expanded from a helpful tool in managing student records to the truth source for any data in the student record, while its databases, containing a huge amount of information, still rely upon a signed paper backed scheme for their data accountability and integrity. This level of integrity assurance is enough for a complementary tool but does not retain enough properties to be itself a truth source which is a requirement to the overall trust in the institution's academic records and therefore to the institution itself.

While some guidelines have been drafted[2] in issuing a Digital Diploma Supplement some extended work in the information system is required to assert the trust in the institutional records themselves, assuring integrity in the data before any diploma is issued and while the students records are being modified inside the information system.

The main objective in this paper is not to describe the information system itself, but to describe our environment as to establish an insight into internal procedures and the associated requirements in providing external entities the mechanism to validate trust in every document issued by the university and retaining the ability to import any document into another information system without human intervention.

2.1. The university structure

The University of Porto has 14 faculties and a business school and provides a large variety of courses while covering the whole range of study areas and all levels of higher education.

The schools of the University have administrative autonomy, this organizational separation creates the need to ensure that the information systems implemented maintain local specificities and allow administrative processes to be dematerialized contemplating the structures defined in the University hierarchy.

An implementation of a technological infrastructure should enable the transfer of University's organizational hierarchies to the digital world, ensuring for example that the generated electronic documents are digitally signed and keeping the required hierarchical dependence.

2.2. Information trust chain

In academic environment, trust has an implied transitive nature with trust delegation methods being in use for student grading and in the internal institution hierarchy.

While trust can be delegated in an individual, it can also be delegated in a set of individuals that can only produce a trusted document when pre-determined conditions are met. Establishing this trust chain requires changes to the information system to cope with the management of trust delegation, defining domains and sets of trust for each task while producing tamper-evident documents describing this delegation chain.

The development of such a trust chain recreates the institutional hierarchy trust which may not be known to an external entity, meaning that every final signed document available outside the information system must contain a signature from the trusted-most signee, namely the institutional signature being provided by the institution's digital signature services and time stamped[3] by an external entity as to offer non-repudiation in every document issued by the University.

2.3. Sigarra database technology

The U.PORTO Information System is supported by a relational database SQL, Oracle technology. Data are added to the database via the application layer of the information system, the information is recorded and simultaneously logged with timestamp values and information about the user who

interacted with the system. All changes to the database are subject to registration through the change logs of DBMS, enabling auditability mechanisms.

2.4. Risks to data integrity

In the current information system, the database has separate administrative domains, but one user with privileges has the possibility to modify any data related to student records, effectively positioning themselves as a trusted party in the institutional trust chain and increasing responsibility and accountability of the staff. Also, any intrusion to the information system or its databases may manipulate information pertaining to student academic information that would be almost blindly trusted by the institution except for a manual verification of the paper trail which, obviously, has associated costs.

When moving away from a paper-backed environment, the threat posed by a lack of trust domains separation is not exclusive to whoever has the ability to manipulate data, but by offering a degree of distrust in the integrity of the data provides the issuer of each grade with the ability to repudiate the previously submitted information, subverting also the required non-repudiation property on a trusted information source.

2.5. Current grading process

Currently at U.PORTO the process of collection and registration of evaluations of students is supported by SIGARRA with various process options being defined in a school by school basis.

The professors of the disciplines are usually responsible for the registration of the student's evaluations in the information system. After the professor finishes all evaluations, the system provides tools to publish and advertise the grades giving students the ability to detect possible errors and request corrections. Once the period of publication ends, the professor locks the evaluations which make the records permanent, losing the temporary character. The evaluations module enables the option to print the document with assessments, allowing the professor to sign and send it to the academic services to archival.

3. DEMATERIALIZATION OF DOCUMENTS

Materialization and dematerialization of documents is a very important process in the University workflow. As described in 2.1 the University of Porto is composed of different faculties that shall be regarded as independent with different workflow rules in administrative processes, requiring reengineering of the documents dematerialization. This nature of independence is akin to what happens in student mobility programs where each party relies on information sharing and the integrity of the information being exchanged.

Verifying a physical document integrity is not a simple process, requiring the verifier to have a pristine trusted copy of the document's signee signature for comparison and trusted documents containing the university stamp. This poses yet another problem, staff training to spot signature forgeries is nonexistent and it is a time consuming task, leading to a huge effort verifying documents in transit between institutions or to the non-verification of their integrity.

Due to the interest in moving to a paper-less environment some alternatives have been drafted to create and implement a framework that could use digital signed documents as a database and archive.

Some guidelines have been drafted in relation to the mobility programs and taken into account during the implementation of our framework's next iteration, namely the compliance with ETSI TS 101 903 (XAdES)[4] or ETSI TS 102 778 (PAdES)[5] which could have an impact not only in the document exchange inside the institution itself but with the partners, facilitating the production of digital documents that could be easily adapted to produce the Diploma Supplements document.

3.1. Digital signed documents

Taking into account the problems identified to ascertain the integrity of physical documents, the need to move to a paper-less environment, and the need for a stronger auditability in the information system, the SQL database was deprecated in favor of an XML database. Leveraging the

possibility to have digitally signed XML documents, each document can be signed by a group of trusted signees, which are described in another set of documents where trust is recursively established from the university top hierarchy to the signer.

This method provides strong accountability and allows the university to easily revoke the delegated trust in each signer, to define different types of documents and corresponding trust domains while allowing the trustee to delegate tasks by signing a document entrusting another signee.

A practical example of applying this trust delegation workflow can be summarized as:

1. The faculty delegates the trust to a department to grade the courses in their faculty.
2. The department delegates the trust of each course class in a set of professors
3. Each professor signs the class grading document. (multiple signatures may be required)
4. When all required signatures are present on the document, the information system submits the document to the institution where it is signed with the institution key and externally time stamped.

This process provides a verifiable chain of signatures and allows the institution to commit to the validity of the delegated trust at the signing time, providing non-repudiation for each signature present in the document. Each document signed with the institutional key is therefore valid and the signers who committed with the information present in the document are accountable.

When a new document needs to be signed, the delegation chain is evaluated to determine who can actually sign a document, being entrusted by the institution. This workflow can be extended to any kind of document produced by the institution and not solely to academic records, establishing a trust anchor point in the trusted most point in hierarchy, the institution itself.

3.2. Electronic graduation documents

Issuing a legal document and committing to the information contained requires the absolute trust in the information source and the described method provides us the ability to verify and assert the integrity of each unit in the student academic achievements without human intervention and subsequently the mechanism to produce institutionally signed documents.

The production of given documents poses yet another challenge, when materializing a document from the information system we can comply with different document types.

While PDF[6] documents are generally accepted as an easily human-verifiable standard for information sharing, the dematerialization of the given document type into an information system isn't straightforward and means the loss of all trusted-chain signatures when importing into a new information system. XML documents on the other hand are easily dematerialized into the information system but do not allow for easy human verification outside the scope of an information system.

A PDF[7] document can be produced with attached XML files that can be used by different information systems. To assure the integrity of both methods both the PDF and the XML documents need to be signed by the institution's private key.

When importing a document with this method, the information system may rely solely on the signed XML document which contains the information used to render the PDF. Human verification is also still possible by verifying the PDF signature.

4. ARCHITECTURE

To provide support for the documents described in 3.1 and maintain the required properties identified in the introduction as to minimize the associated risks, the entire infrastructure of the information system had to be redesigned as described in Figure 1.

A Java applet was developed in the information system providing the support for digitally signing XML documents generated by the information system itself. Signing such a document binds the user to the information contained in the document offering non-repudiation to the information system.

Two-party non-repudiation is required when assuring to each signer that the data is in fact accepted by the institution and archived accordingly, this process requires an institutional signature to be produced as to bind the institution to the reception of documents and a timestamp obtained from an

DOI: 10.7250/eunis.2013.040

external entity. Supporting this institutional signature and timestamp requires a SignServer validating the document structure, signing and archiving documents.

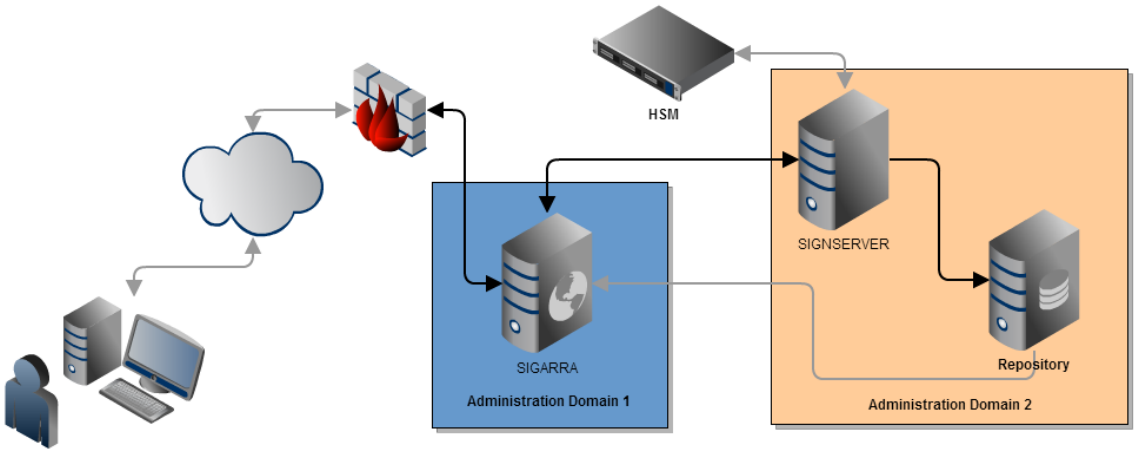

Figure 1. Supporting Architecture

Using the information system, multiple users use their tokens to digitally sign the XML documents in the information system, which ending this collection sends them to the institution signature service, supported by PrimeKey's SignServer[8]. Since all documents present in the repository are required to be signed and must not be manipulated at will by the information system, two domains are defined in the repository. The SignServer has write privileges to the repository to store signed documents but the information system can only read these documents while having transitive write privileges by using the SignServer to verify signature validity, sign and write them to the database.

The SignServer is maintained by the staff, but by shifting all the trust to a single service, the set of responsible and accountable personnel is effectively reduced, and by using a Hardware Security Module to store the private keys we assert that only a group of coordinated personnel can manage the service private keys.

5. CONCLUSION

The dematerialization of University documents and its integration with the information system required the shift to a new document format facilitating information sharing across different trust domains.

This document should not only comply with certain properties as to maintain integrity mechanisms and strong non-repudiation but must provide verification methods both automatic and by hand for document materialization and dematerialization into different information system domains.

To fulfill the requirements we used XADES signed XML files, implementing a central repository containing not only academic records but also the trust delegation chains.

By leveraging the PDF's specification, namely signed dynamic XFA forms, the production of a versatile document, independent of information system and with offline verification by commonly present tools is possible, effectively enabling the production of documents that meet the intended requirements.

6. REFERENCES

Lígia M. Ribeiro, G.D., Ana Azevedo and J. C. Marques dos Santos. (1997) Developing An Information System At The Engineering Faculty Of Porto University. in EUNIS. Grenoble, France.

Grant, S. (2009) Guidelines on a European Learner Mobility model. 15 January 2013]; Available from: http://wiki.teria.no/display/EuropeanLearnerMobility/Guidelines+on+European+Learner+Mobility.

Council, E.P.a. (1999), Community framework for electronic signatures., in Directive 1999/93/EC.

Institute, E.T.S. (2002), XML Advanced Electronic Signatures (XAdES), ETSI TS 101 903 version 1.4.1,.

DOI: 10.7250/eunis.2013.040

Institute, E.T.S. (2009), Electronic Signatures and Infrastructures (ESI); PDF Advanced Electronic Signature Profiles; Part 5: PAdES for XML Content - Profiles for XAdES signatures..

Systems, A. (2001), PDF Reference, in Adobe Portable Document Format, ADDISON–WESLEY, Editor.

Systems, A. (2009), The AdES family of standards: CAdES, XAdES, and PAdES Implementation guidance for using electronic signatures in the European Union..

Vendil, P. (2009), SignServer Manual, Primekey, Editor.

7. AUTHORS' BIOGRAPHIES

Luís A. Maia is currently an MsC student of Networking Engineering at the Science Faculty of Porto University and also a member of Instituto de Telecomunicações, where he holds a research position. His main focus is computer and network security, Mobile Systems and cryptography. For the period of two years he studied at the International Faculty of Engineering (Lodz) before joining the CRACS-INESC team and at a later stage – the current institution.

Luís M. Valente has an MsC in Network and Information Systems Engineering and currently works on implementation of Electronic Administration projects, integrated in Digital University Department of the U.PORTO. He is also a researcher in the field of computer security at the CRACS/INESC-LA Port.

Working straight in the U.PORTO Campus Card Project, his main assignment is the implementation of cryptographic functionalities, especially for secure authentication and digital signing.

Manuel E. Correia got his MSc in foundations of advanced information processing technologies from the Imperial College of London in 1992 and his PhD in Computer Science from Oporto University in 2001. He is currently a lecturer at the Department of Computer Science of the Faculty of Science of Oporto University and a researcher in the field of computer security at the CRACS/INESC-LA Porto, where he is responsible for research projects related with user centric digital identity management and token based authentication and the information security aspects of several industry contracts. He is also a consultant for some Portuguese public agencies (Health and Education) in computer security.

Lígia M. Ribeiro received her degree in Applied Mathematics in 1977 at the University of Porto and holds a PhD in Engineering Science from the University of Minho.

She is pro-rector at the University of Porto, being responsible for ICT. She is also a Principal Researcher at the Faculty of Engineering of the University of Porto, where she was director of the Computer Centre between 1997 and 2002. Lígia Ribeiro lectured at the Faculty of Sciences of the University of Porto between 1978 and 1988, being Assistant Professor within the group of Applied Mathematics. In 1988 she started to work at the Engineering Faculty of the same University as a researcher. She was vice-president of the Institute for Common Resources and Initiatives of the University of Porto between 2003 and 2009.

DOI: 10.7250/eunis.2013.040

Lígia Ribeiro was President of EUNIS, the European University Information Systems Organization, between 2004 and 2006, after being vice-president for two years. She is presently a member of the EUNIS Board of Directors. She was also a member of the Technical Committee of TERENA, the Trans-European Research and Education Networking Association, between 2009 and 2011. She is a member of ACM, the Association for Computing Machinery.
Her main research areas of interest are Computer Simulation, Distributed Computing and High Performance Computing, and Information Systems. Lígia Ribeiro was co-responsible for the development of the Information System of the University of Porto (SIGARRA) and she is nowadays the coordinator general of this IS.

Luís Antunes obtained a PhD in Computer Science at University of Porto. Currently he is an Associated Professor at the Computer Science Department at the University of Porto. Most of his research is on Computational Complexity and Cryptography. He is a co-author of more than 30 indexed publications and the PI of some research projects founded by the Portuguese National Science Foundation. He supervised several Master and PhD students in the past and currently supervises two PhD students and several Master students in areas such as Access Control and Information Measures for Cryptography Protocols.
He founded and directed the first Health Informatics Master course in Portugal. He is also a consultant for some Portuguese public agencies (Health and Education) in computer security.