



Grant Agreement no: 262448

Research for SMEs (FP7-SME-2010-1)

Final publishable report

28-01-2013

Start date of the Project: 1st December 2010

Duration: 24 months

Coordinator: Bit Oceans Research S.L. (BIT)

Version: 1.0

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 262448. The information in this document is provided “as is” and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

CONTENTS

1. EXECUTIVE SUMMARY	1
2. SUMMARY DESCRIPTION OF PROJECT CONTEXT AND OBJECTIVES	2
3. DESCRIPTION OF THE MAIN S&T RESULTS	5
3.1. RESULT 1: ROBUST HASH MODULE.....	5
3.2. CUSTOMISABLE SECURITY MODULE	6
3.3. CRYPTOGRAPHIC TOOLS.....	7
3.4. PRINT&SCAN CHANNEL MODELS	7
3.5. ESTIMATORS OF CHANNEL MODEL PARAMETERS	8
3.6. TOOL FOR CORRECTING PRINT&SCAN DISTORTIONS	8
3.7. SYNCHRONISATION MODULE	9
3.8. DOCUMENT COMPARISON MODULE	10
3.9. USER INTERFACES.....	12
3.10. SYSTEM ARCHITECTURE	16
3.11. SECURITY TESTING METHODOLOGY	17
3.12. THE SIGNED PROTOTYPE.....	17
3.13. MAIN PERFORMANCE INDICATORS.....	18
3.14. CONCLUSIONS	19
4. POTENTIAL IMPACT	21
4.1. DISSEMINATION AND EXPLOITATION ACTIVITIES	26
5. ANNEX I – ACRONYMS AND ABBREVIATIONS	28
6. ANNEX II – CONTACT SHEET	29

LIST OF FIGURES

FIGURE 1: SIMPLIFIED WORKFLOW OF THE SIGNED SECURISATION PROCESS	4
FIGURE 2: SIMPLIFIED WORKFLOW OF THE SIGNED VERIFICATION PROCESS.....	4
FIGURE 3: PROBABILITY OF COLLISION FOR THE HDWT FUNCTION APPLIED ON DOCUMENTS SECURISED AND SCANNED AT 150 DPI (LOW SECURITY LEVEL)	6
FIGURE 4: SIGNAL TO NOISE RATIO GAIN AFTER EQUALIZATION.....	9
FIGURE 5: HISTOGRAM OF THE MISALIGNMENT (IN PIXELS) BETWEEN THE BLOCKS OF THE ORIGINAL DOCUMENT AND THOSE OF THE SCANNED DOCUMENT AFTER SYNCHRONIZATION.....	10
FIGURE 6: PROBABILITIES OF FALSE ALARM (PFA) AND MISSED DETECTION (PMD) FOR A) REPLACEMENTS OF DIGITS (3 BY 8) IN ARIAL 8 FONT; B) REPLACEMENT OF DOTS BY COMMAS IN VERDANA 10 FONT	11
FIGURE 7: FALSE ALARM VS. FALSE REJECTION RATE FOR PHOTOCOPY DETECTION	12
FIGURE 8: HOME SCREEN OF THE SIGNED WEB APPLICATION	13
FIGURE 9: CONTEXT MENU OF THE SIGNED WEB APPLICATION	13
FIGURE 10: SECURISATION DIALOG SHOWING A RECENTLY SECURISED DOCUMENT.....	14
FIGURE 11: EXAMPLE OF A VERIFIED TEXT DOCUMENT, GENERATED BY THE SIGNED PROTOTYPE. NOTICE THE HIGHLIGHTED REGIONS, POINTING OUT THE SIX DETECTED MANIPULATIONS (FOUR REPLACEMENTS OF COMMAS BY DOTS, AND TWO REPLACEMENTS OF DIGITS).....	15
FIGURE 12: EXAMPLE OF A VERIFIED GRAPHIC DOCUMENT, GENERATED BY THE SIGNED PROTOTYPE. THE MANIPULATED REGIONS ARE HIGHLIGHTED.	16

1. Executive summary

In recent years, the concerns about forged documents have been increasing. With the availability of sophisticated yet easy to use tools for manipulating digital images, nowadays it is very simple to manipulate any paper-based document by digitalising it (obtaining an image through scan), manipulating the image through image editors, and then printing it. The manipulated document will look authentic to users, and no easy-to-use systems are available to verify the authenticity of such document. Nowadays, the most widely extended method for checking the validity of printed documents is the mere visual inspection of the document to be validated and a trusted copy. Besides being largely prone to human errors and highly inefficient, manual validation is not always possible because the original trusted document is not always available. Certain sensitive documents, like IDs, are protected through physical-chemical-optical security measures, but their validation must be performed with the help of special hardware devices, and in general the application of certain experts' personal criteria to do the final decision about their authenticity.

There is the need, therefore, of a system able to authenticate paper based documents, recognising if a certain paper document has been manipulated after being issued. The SIGNED project faces this need, providing a trustable information exchange through paper medium which permits to avoid fraud and deception with printed documents. The final objective of the SIGNED solution is to extend to paper-based documents the same degree of trust in authentication, and the same security level, as the well-established (for digital documents) Digital Signature techniques, providing the following guarantees to the users:

- **Integrity:** the receiver of the document is always able to determine if any change in the document, introduced after the document has been signed, occurred.
- **Authenticity:** the receiver of the document is always able to reliably identify the issuer of the document.
- **Non-repudiation:** the issuer cannot deny the production of the signed document.

To achieve such level of security on paper based documents, the solution to be developed in the SIGNED project addresses a set of functional requirements and technical issues to be overcome. The most important are the following:

- It must be able to convey security information along with the document itself, since the paper is the only available transport means.
- With a negligible probability of not detecting a fraudulent manipulation, and clearly spotting the regions of the document that have been manipulated. It must be able to detect changes that are meaningful to the reader (semantic changes) and ignore those distortions introduced during the printing and scanning processes (accidental changes).
- Applicable to any kind of document, regardless of its format and layout.
- Compatible with the widest possible range of printers and scanners, including fax machines if possible.
- Highly efficient, from the computational point of view, usable by the general public (i.e. non-specialists in fraud detection) and easily integrable into the users' workflows.

2. Summary description of project context and objectives

The objective of the SIGNED project is to develop a paper-based document authentication system, making use of innovative image hashing techniques, integrated into a flexible tool usable in a standalone way or integrated with organisations' systems.

Nowadays, the most widely extended method for checking the validity of printed documents is the mere visual inspection of the document to be validated and a trusted copy. This is a highly inefficient process, since it consumes a lot of time and human resources. Moreover, this manual validation is not always possible because original trusted document is not available, and it is largely prone to human errors, especially when a large volume of documents must be validated in a short time. Usually, the authenticity of a printed document is taken for granted upon the existence of a hand-written signature or some sort of stamp. In certain documents like IDs or academic degrees, physical-chemical-optical security measures are applied such as fluorescent inks, watermarks and holograms. However, in order to perform the validation of one of these documents, it is necessary the availability of special (and expensive) hardware devices, and in general the application of certain experts' personal criteria to do the final decision about authenticity.

The impact of forged documents has been increasing in recent years, as far as it is connected with the availability of sophisticated yet easy to use tools for manipulating digital images: nowadays, in fact, a paper based document could be easily manipulated by digitalising it (obtaining an image through scan), manipulating the image through image editors, and then printing it; the printed document will look authentic to users, and no easy-to-use systems are available to verify the authenticity of such document.

There is the need, therefore, of a system able to authenticate paper based documents, recognising if a certain paper document has been manipulated after being issued. The SIGNED project indeed faces this need, providing a trustable information exchange through paper medium which permits to avoid fraud and deception with printed documents. The project results enable SMEs, Public Administration, Banks, Insurance Companies and citizens to verify the authenticity of sensitive paper based documents, such as financial information, personal identity documents, etc.

The objective of the SIGNED solution is to extend to paper-based documents the same degree of trust in authentication, and the same security level, as the well-established (for digital documents) Digital Signature techniques, providing the following guarantees to the users:

- **Integrity:** the receiver of the document is always able to determine if any change in the document, introduced after the document has been signed, occurred.
- **Authenticity:** the receiver of the document is always able to reliably identify the issuer of the document.
- **Non-repudiation:** the issuer cannot deny the production of the signed document.

To achieve such level of security on paper based documents, the SIGNED project addresses a set of particular objectives in the form of functional requirements and technical issues to be overcome, in order to obtain an integrity, authenticity and non-repudiation property level

similar to those of the DSAS. The main functional requirements of the document authentication solution are the following:

- It must be able to convey security information along with the document itself, since the paper is the only available transport means.
- It must be applicable to any kind of document, regardless of its format and layout (this is analogous to the digital signature, which is applicable to any kind of bit stream, no matter its semantic content).
- Likewise, the SIGNED solution must be compatible with the widest possible range of printers and scanners, including fax machines if possible.
- It must have a negligible probability of not detecting a fraudulent manipulation. As a quantitative objective, the probability of detecting the replacement of a dot by a comma in 10 points font must be above or equal to 99.9%.
- It must clearly spot the regions of the document that have been manipulated (i.e. it should not spot the whole document as a forgery if only one small region has been tampered with).
- It must have a high computational efficiency in order to allow for a good throughput which does not prevent its application in demanding real-time scenarios, as it could be the financial one. As a quantitative objective, the time needed for securing a document using an average process server must be below two seconds.
- It must be able to detect changes that are meaningful to the reader (semantic changes) and ignore those distortions introduced during the printing and scanning processes (accidental changes).
- It must be usable by the general public (i.e. non-specialists in fraud detection) and easily integrable into the users' workflows.

To obtain an authentication scheme that can comply with the above described functionalities, the SIGNED project pursues the following main S&T objectives:

1. Formulation of a suitable overall system architecture, capable of fulfilling all the requirements.
2. Formulation and development of a new hashing algorithm robust to the noise introduced by the Print&Scan channel, able to distinguish meaningful changes (i.e. manipulations) from accidental ones (due to the Print&Scan process), with a compact digest, and applicable to any kind of printed document (i.e. independent of its layout).
3. Selection of a method, based on barcode technology, to embed security information in a document using standard printing hardware.
4. Formulation of accurate models of the digital-to-analog and analog-to-digital conversion processes. Use of these models for designing methods that can mitigate the distortions introduced by the Print&Scan channel.
5. Formulation of methods for providing an additional level of understanding to the manipulations undergone by a document, determining for instance whether a given document has undergone photocopying or double scanning, or whether the identified manipulations are due to accidental changes such as scratches and folding.

The project final outcome is a new authentication system for documents (both in digital and in paper format), based on hashing technologies, and incorporated in an easy to use IT tool (to be

used in a standalone modality or to be integrated in corporate systems). The developed system is able to provide a unique security seal for any paper based document, allowing to detect if changes occurred to the document after the security seal was inserted.

The workflow of the SIGNED securisation and verification processes are illustrated in Figure 1 and Figure 2.

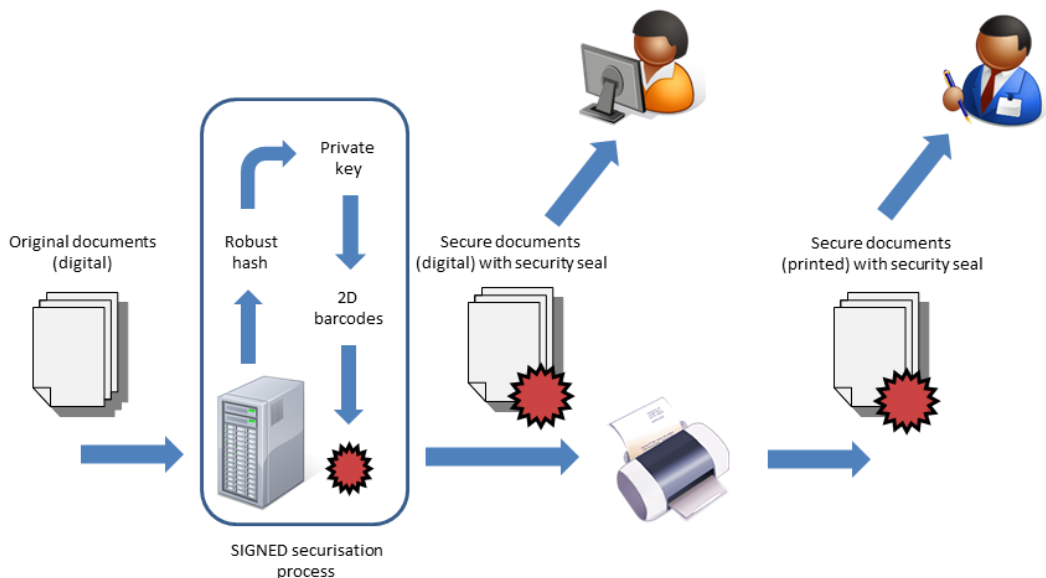


Figure 1: simplified workflow of the SIGNED securisation process

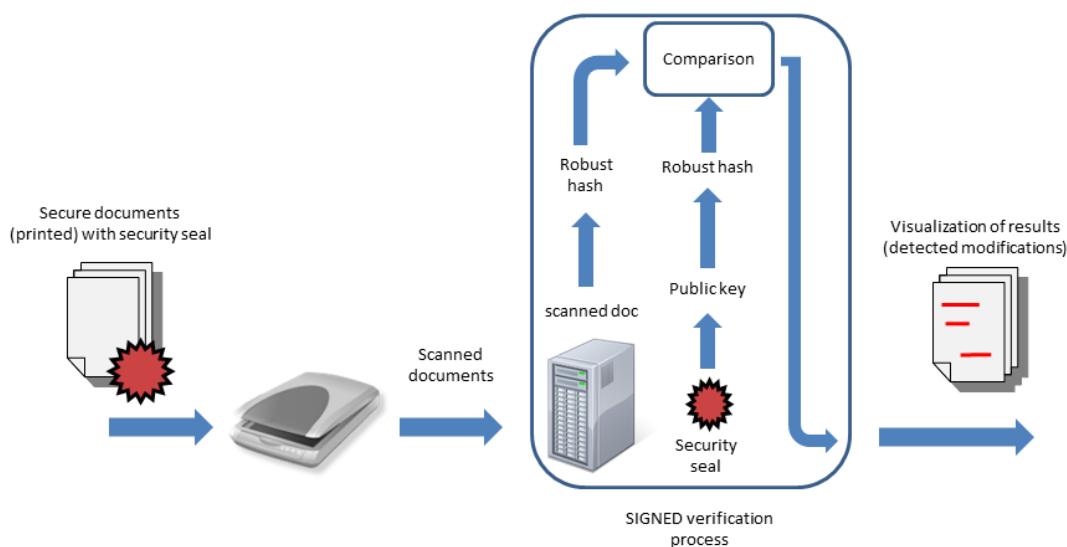


Figure 2: simplified workflow of the SIGNED verification process

3. Description of the main S&T results

The S&T results of SIGNED fulfil the five S&T objectives established at the beginning of the project (see previous section). The main results, along with their mapping to the relevant objectives, are enumerated below.

- **Result 1:** Robust hash module (software tool). Related objectives: 2, 3
- **Result 2:** Customisable security module (software tool). Related objectives: 2
- **Result 3:** Cryptographic tools (software tool). Related objectives: 1
- **Result 4:** Print&Scan channel models (specification). Related objectives: 4
- **Result 5:** Estimators of channel model parameters (software tool). Related objectives: 4
- **Result 6:** Tool for correcting Print&Scan distortions (software tool). Related objectives: 4
- **Result 7:** Synchronisation module (software tool). Related objectives: 4
- **Result 8:** Document comparison module (software tool). Related objectives: 4, 5
- **Result 9:** User interfaces (software tools). Related objectives: 1
- **Result 10:** System architecture (specification). Related objectives: 1
- **Result 11:** Security testing methodology (specification). Related objectives: all (as this result deals with the validation of all previous results)

The results 1, 2, 3, 9 enumerated above, properly integrated, form the “secure document generation module” of the SIGNED solution. Likewise, the integration of the results 1, 3, 4, 5, 6, 7, 8, 9 forms the “secure document verification module”. The integration of both modules is the SIGNED prototype, which constitutes the major result of the project, as it is a complete software demonstrator comprising all the functionalities of the SIGNED solution.

The SIGNED prototype and the S&T results are further described in the following.

3.1. Result 1: robust hash module

The first major result of SIGNED is the specification and software implementation of a robust hash module that allows to represent digital documents in a manner which is robust to the distortions introduced by the Print&Scan channel. Such module is applicable to all kind of digital documents, independently of their layout and format, as it treats the documents as digital images. Furthermore, it is applicable to B&W, grayscale and color documents, although it was optimized for working with grayscale text documents, which is the scenario of highest interest for the SMEs participating in the project.

The core of the robust hash module is a new robust hashing function, named HDWT, which was designed specifically to meet the requirements of the project, namely: a hash length as short as possible, and a high discriminability for detecting modifications. Figure 3 illustrates the probability of false alarm (PFA) vs. the probability of collision (Pcollision) obtained for the HDWT function, showing that it is possible to obtain PFA and Pcollision below 0.001. The HDWT hashing function itself is a valuable outcome of the project, as it has wide applicability in the image identification and authentication fields.

The robust hash module also comprises an auxiliary module which is responsible for embedding the security information (e.g. the robust hash) in the securised documents, by means of off-the-shelf, open-source QR (Quick Response) barcode technology whose parameters were optimized for an optimal performance in the SIGNED solution. In addition, a higher processing layer for the information embedded in the barcodes was added in order to improve the robustness of the selected barcode technology.

The result 1 is part of the “secure document generation” and “secure document verification” modules of the SIGNED solution, and it must be used in conjunction with results 4, 5, 6, and 7 in order to fully address the requirements established at the beginning of the project.

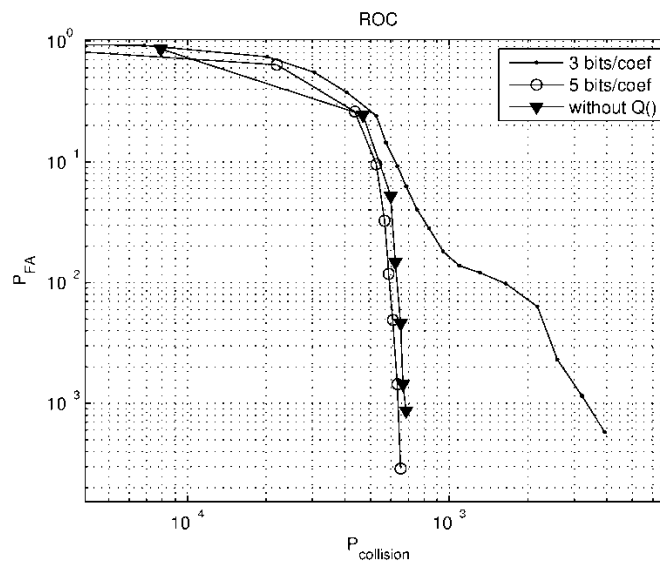


Figure 3: probability of collision for the HDWT function applied on documents securised and scanned at 150 dpi (low security level)

3.2. Customisable security module

The result 2 is part of the “secure document generation” module of the SIGNED solution. The aim of the customisable security module is to trade-off the behaviour of the SIGNED solution among different performance criteria, which allows to adapt the SIGNED solution to different scenarios of interest. Those indicators comprise, for instance, the detection capabilities and the throughput of the securisation/verification processes. The trade-off is achieved by properly tuning the parameters of the HDWT hashing technique, in a way that would allow to apply different security levels for different regions of the document. This last feature is interesting, for instance, in the case of personal IDs where the authenticity of the picture and the name of the ID holder are of extreme importance.

3.3. Cryptographic tools

Recall that the main objective of the SIGNED solution is to extend to paper based documents the concept of Digital Signature (DSAS) technique, providing integrity, authenticity and non-repudiation. In order to fulfil these requirements the SIGNED solution incorporates a cryptographic layer, which relies on the use of digital certificates and digital signatures. Digital certificates link the identity of the producer of the securised document (the signer) to the document itself, by digitally signing the information embedded in the barcodes of the securised document. In this manner, the SIGNED solution guarantees authenticity (the receiver of the document is always able to reliably identify the signer) and non-repudiation (the signer cannot deny the production of the signed document). Furthermore, the digital signature guarantees that the receiver of the document is always able to reliably determine if any change in the document, introduced after the document has been signed, occurred. This is so because any modification to the information embedded in the barcodes of the securised document could be detected.

In addition, the cryptographic layer of the SIGNED solution includes a set of protocols to guarantee the secure exchange of information between the signer and the secure document generation module, independently of the location of the latter. Thanks to these protocols, the SIGNED solution is usable both in local and remote deployments, supporting SaaS commercialization services.

3.4. Print&Scan channel models

The main innovation of the SIGNED solution is its capability to reliably provide authentication of documents even in printed format, thus extending the traditional concept of digital signature to paper documents. As the SIGNED solution is intended to work with any type of document, standard paper, and standard printing and scanning hardware, it is easy to understand that such innovation actually represents the major technological challenge of the project.

The Print&Scan channel models of the SIGNED solution are the mathematical specification of the distortions suffered by a digital document which is printed and scanned. The distortions suffered by a document undergoing printing and scanning are numerous and highly dependent on the printing and scanning devices. Thus, a good modelling of those distortions was crucial to the success of the project. The Print&Scan channel model that was designed in the project is generic enough to model the wide range of existing devices, but at the same time specific enough to achieve the necessary degree of accuracy for detecting tiny malicious alterations of the documents.

The type of distortions addressed by the Print&Scan channel model can be summarized in two:

1. **Geometric distortions.** The Print&Scan channel suffers from certain geometric transformations which desynchronize (i.e. misalign) the scanned document with the original one. The geometric distortions introduced by the Print&Scan channel are very

strong and need to be corrected (or at least mitigated at a great extent) in order to ensure the good performance of the verification process.

2. **Amplitude distortions.** Besides the geometric distortions, the P&S channel introduces distortions that modify the values of the pixels of the scanned documents. The robust hash function must provide a certain degree of robustness to such distortions. However, the performance of the verification process can be significantly improved by (at least partially) inverting the amplitude distortions.

The SIGNED channel model addresses directly the distortions suffered by the robust hash coefficients, thus it could be adapted in the future to possible evolutions of the robust hashing function. Moreover, the model is applicable both to grayscale and color documents.

The results 5, 6, 7, and 8 build upon the Print&Scan channel model. Thanks to the availability of such model, the distortions introduced by the Print&Scan channel can be mitigated and in some cases completely removed, as will be explained below.

3.5. Estimators of channel model parameters

The parameters of the Print&Scan channel model usually have a large dependence on the particular printing and scanning devices, so a generic model does not allow to meet the requirements set for the SIGNED solution. Thus, it is necessary to design and implement techniques that can estimate the Print&Scan channel parameters whenever a securised document must be verified. Result 5 is the specification and software implementation of the channel estimation methods that allow to estimate the relevant parameters. One of the main features of the designed estimators is that they do not require neither the original document in the verification stage, nor the knowledge of the printer and scanner involved in the processing of the document under verification.

The channel estimators of the SIGNED solution succeed in reliably estimating the parameters of both the geometrical and the amplitude distortions envisaged by the Print&Scan channel model (result 4). Their good performance, together with results 6 and 7, was confirmed by tests carried out over a large amount of real documents. Moreover, they have a low computational complexity.

The output produced by the result 5 is used by the results 6 and 7 to correct the Print&Scan distortions.

3.6. Tool for correcting Print&Scan distortions

The result 6 consists in the specification and software implementation of tools to correct the distortions of the Print&Scan channel, using the estimates of the channel parameters provided by result 5. These tools, also named “equalizers”, focus on correcting the amplitude distortions

envisaged by the Print&Scan channel model (cf. result 4). The aim of the equalizer is to provide a stable representation of the robust hash which is as independent as possible of the particular devices involved in the printing and scanning process.

The equalizer of the SIGNED solution is adaptive, in the sense that it automatically adapts to the document at hand, without the need to know the printer/scanner model that produced the document to be verified. It does not even need a previous training stage. The role of the equalizer is crucial to ensure good performance of the SIGNED solution.

The equalizer is applied on the output of the synchronization module (cf. result 7), whereas the output of the equalization process is directly used by the document comparison module (cf. result 8). The good performance of the equalizer was confirmed by tests carried out over a large amount of real documents. Figure 4 illustrates the gain, in signal to noise ratio, obtained for each component of the HDWT hash in a typical scanned document: as can be seen, the gain is up to 6 dB for several components. Moreover, the equalization process requires a very low computational complexity.

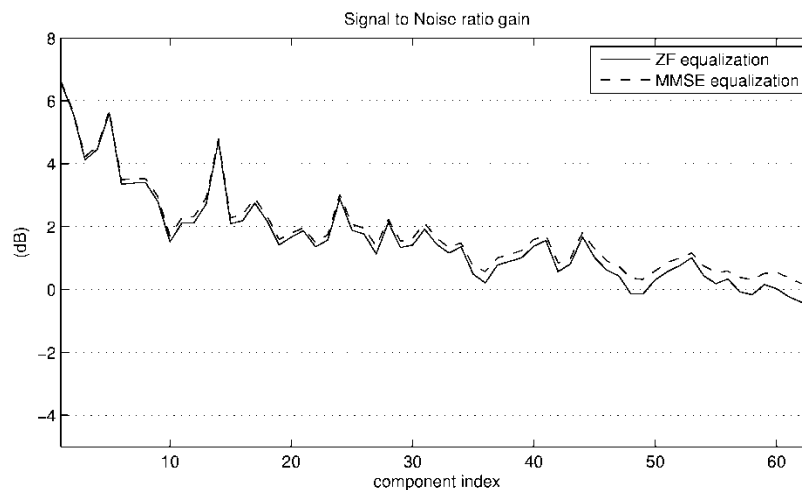


Figure 4: signal to noise ratio gain after equalization

3.7. Synchronisation module

The result 7 consists in the specification and software implementation of tools to correct the geometric distortions introduced by the Print&Scan channel. Right after extracting the information from the embedded barcodes, the synchronization is the first process applied by the SIGNED document verification module. Its task is to “align” as accurately as possible the scanned document to the original document, even if this is not available in the document verification phase. Hence, it can be seen as a sort of “image registration” module. The synchronization module is crucial to ensure the good performance of the equalizer (cf. result 6) and the document comparison module (8). The synchronization module is truly one of the keys to the success of the whole SIGNED solution.

The synchronization module of SIGNED performs two different tasks: gross synchronization and fine synchronization. The first task provides a gross geometrical correction of the scanned document, and the second task does the fine adjustment needed to align the scanned and original documents at a pixel level, which is absolutely necessary in order to provide accurate tamper localization and the low error probabilities required.

The synchronization module of the SIGNED solution has been thoroughly optimized in terms of computational performance, as it is the most time-consuming part of the SIGNED document verification module. In fact, the synchronization process is performed in two steps (gross synch and fine synch) in order to alleviate the computational burden of the process.

The synchronization module finally implemented in the SIGNED solution succeeds in reliably inverting the geometrical distortions of the Print&Scan channel. Its performance was thoroughly tested in a large database of real documents. An example of the results is illustrated in Figure 5.

The synchronized documents are the input to result 6.

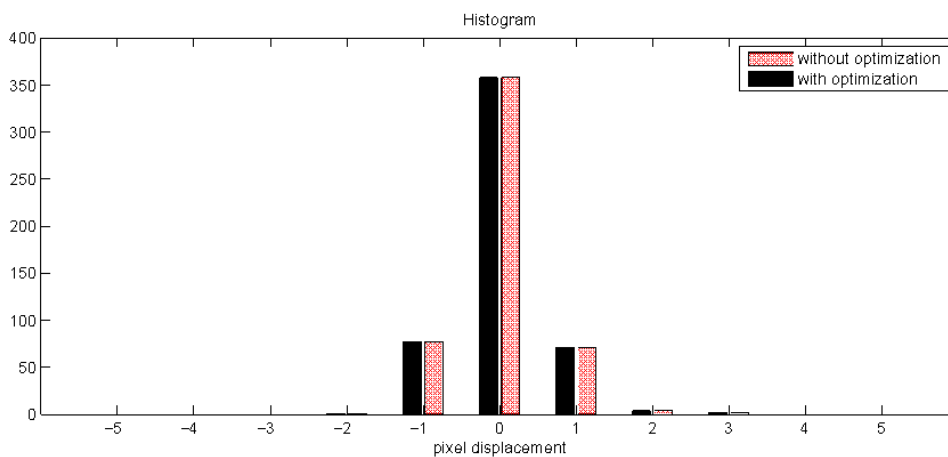
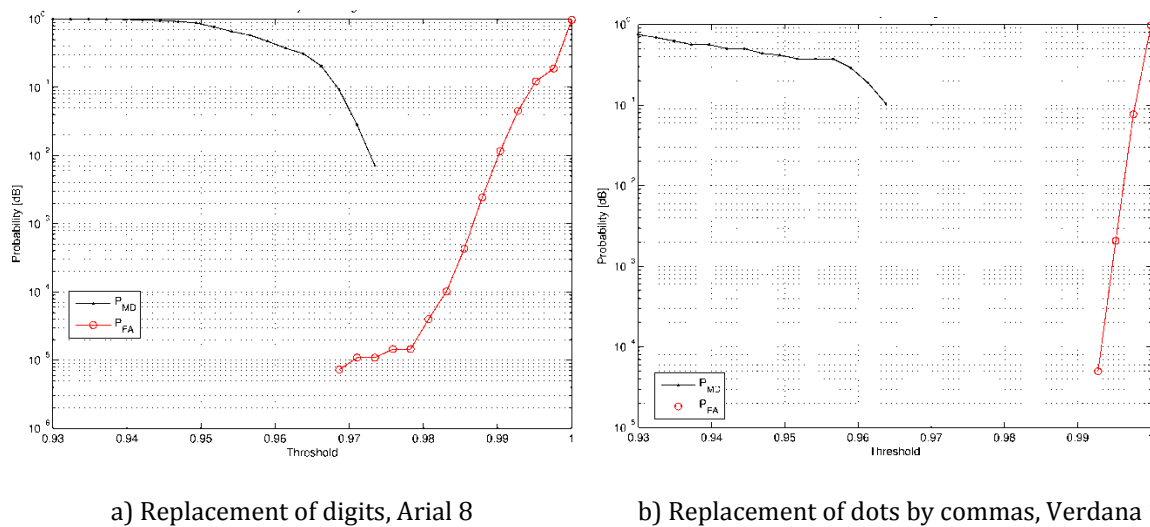


Figure 5: histogram of the misalignment (in pixels) between the blocks of the original document and those of the scanned document after synchronization

3.8. Document comparison module

The result 8 consists in the specification and software implementation of the methods for taking the decision about the authenticity of a document and spotting the alterations to its content, if they exist. The document comparison module fulfils two functional requirements: detecting intentional manipulations and detecting double scanning (photocopying) of the document under verification. The second functionality is actually independent of the securisation process, as it also works with documents which have not been previously securised with the SIGNED solution. It is a functionality required in certain scenarios where it is necessary to distinguish between original documents and copies.

For the detection of intentional manipulations, the document comparison module consists of two steps: first, reading of the information contained in the barcodes of the document to be verified, and second, comparing the output of the equalizer (cf. result 6) to the hash of the original document to identify the possible alterations. Such comparison is not a mere bit to bit comparison, as it comprises an appropriate “distance” computation between hashes and the application of a “smart layer” of processing which takes the final decision of whether a certain region of the document has been altered or not. Of course, this comparison is not affected by the normal distortions of the Print&Scan channel, as it was primarily required for the SIGNED solution.



a) Replacement of digits, Arial 8

b) Replacement of dots by commas, Verdana 10

Figure 6: Probabilities of false alarm (PFA) and missed detection (PMD) for a) replacements of digits (3 by 8) in Arial 8 font; b) replacement of dots by commas in Verdana 10 font

The document comparison module makes use of the digital signature functionality provided by result 3 in order to check the validity of the information extracted from the barcodes embedded in the securised document. If such information is not authentic it is a hint that it was modified by a malicious user, thus the original hash cannot be trusted for performing the comparison and the whole document must be deemed falsified.

The detection of photocopying and double scanning works in a way completely different to the detector of manipulations. In a previous training stage, the detector “learns” the main differences between an original document and a photocopied or double-scanned version. These differences are not extracted directly from the pixels of the document, but extracted from features derived from them. When a document is input to the detector, a smart processing layer takes the decision about its originality based on the learnt features and certain decision rules. Figure 7 illustrates the performance of this tool in a real set-up, with “false alarm” meaning that an original document is deemed photocopied, and with “false rejection” meaning the opposite.

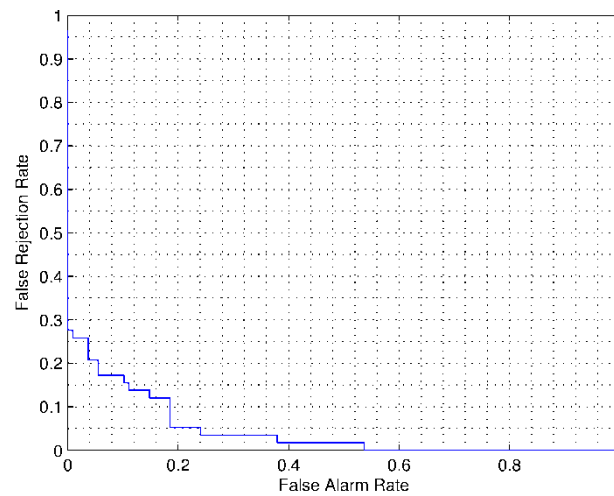


Figure 7: False alarm vs. false rejection rate for photocopy detection

3.9. User interfaces

For maximizing the usability and future integrability of the system, the interfaces to both the securise document generation module and the secure document verification module of SIGNED were implemented as:

- **Web applications** with a common interface, which can be accessed from any web browser and operating system without the need for the user of installing any additional components. This interface makes use of standard web protocols, ensuring simplicity of maintenance and evolution.
- **Web services** for allowing the integration of SIGNED in other external applications. This interface also makes uses of standard protocols (e.g. SOAP) for ensuring its compatibility.

The interface to the secure document generation module comprises the selection of the document to be securised and the interface to the customisable security module, which allows the selection of very high, high and medium security level securisation. The selection of the user is automatically linked by the SIGNED solution to the parameters of the robust hash optimized for each level of security.

The interface to the secure document verification module comprises the selection of the document to be verified (previously securised with the SIGNED solution) and the verification of photocopy/original document. No other input is requested to the user.

Both interfaces (web application and web service) are implemented with standard technologies, ensuring full compatibility.

Figure 8, Figure 9 and Figure 10 below illustrate with several screen captures the interface of the SIGNED web application, showing the processes of document securisation and verification. Figure 11 and Figure 12 show a couple of verified documents, generated by the SIGNED

prototype after running the verification process on documents that have undergone malicious manipulations, printing and scanning.

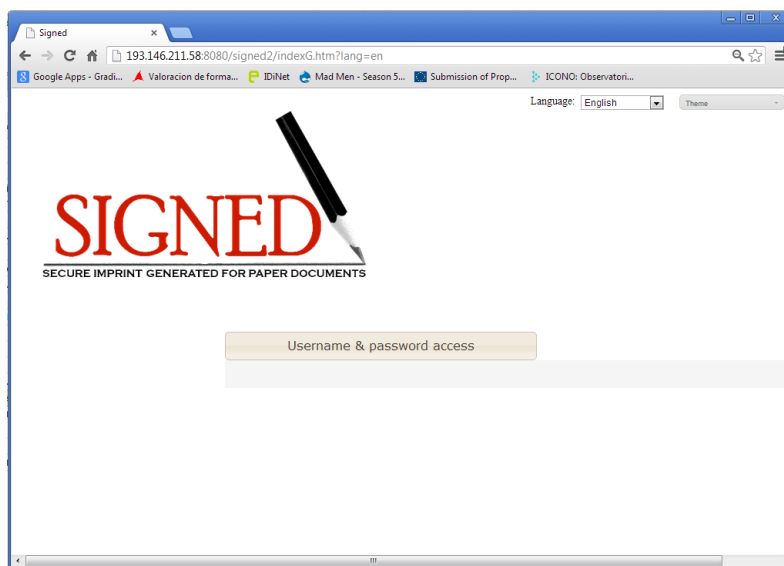


Figure 8: home screen of the SIGNED web application

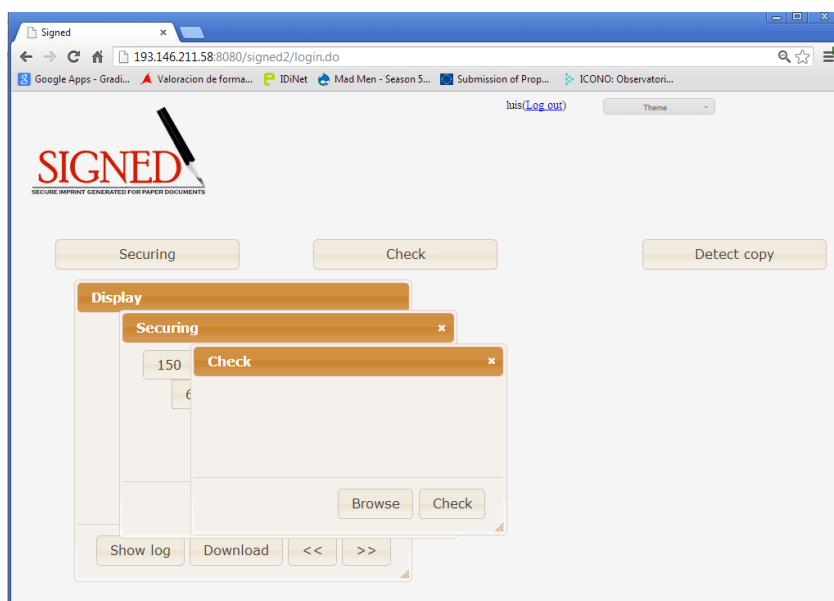


Figure 9: Context menu of the SIGNED web application

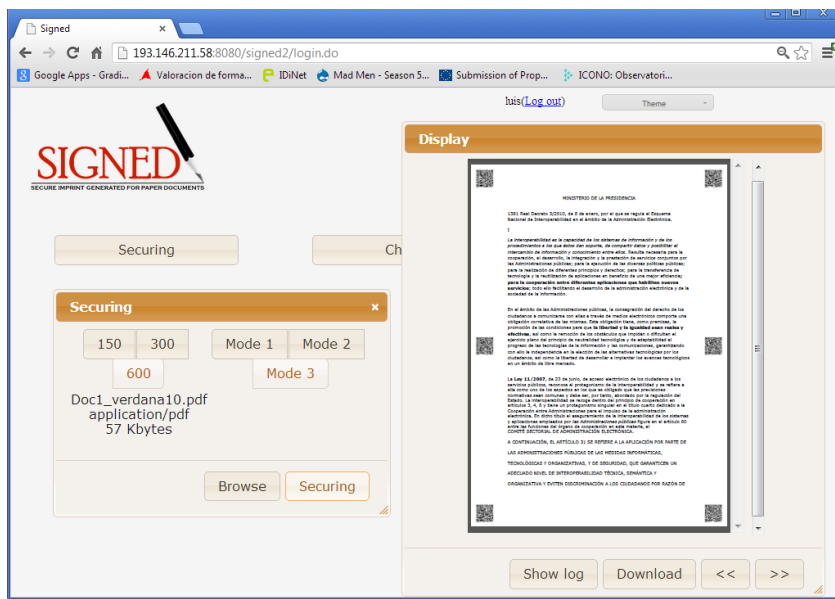


Figure 10: Securisation dialog showing a recently securised document

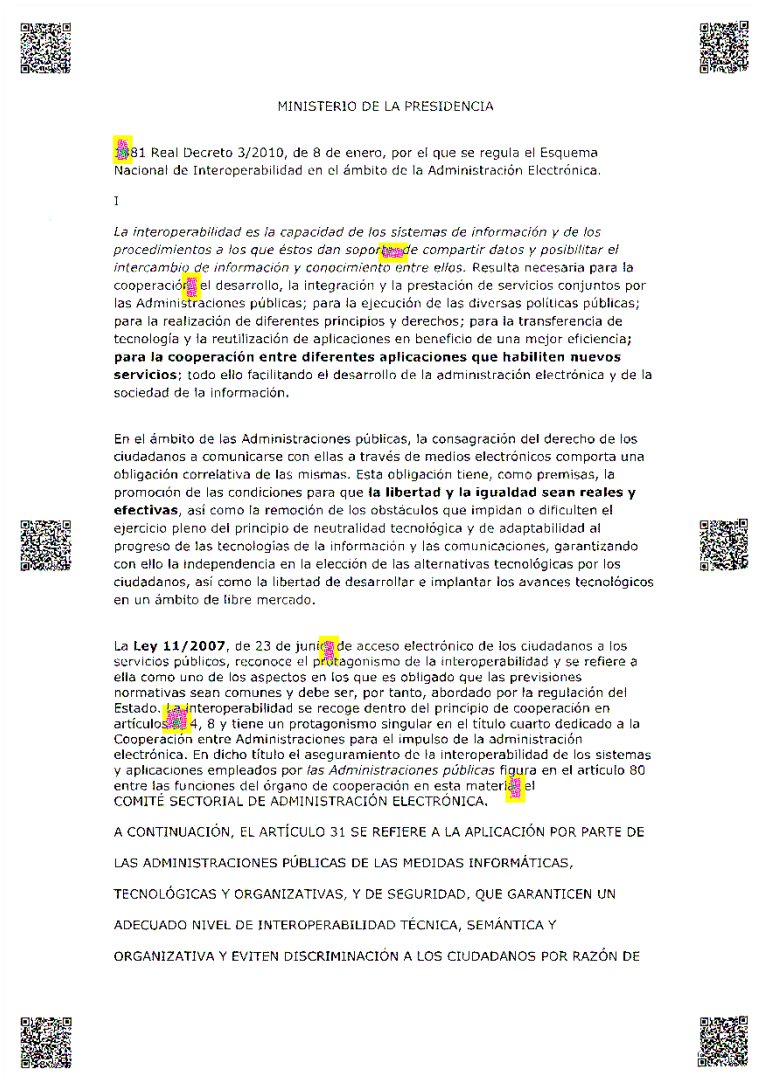


Figure 11: example of a verified text document, generated by the SIGNED prototype. Notice the highlighted regions, pointing out the six detected manipulations (four replacements of commas by dots, and two replacements of digits)

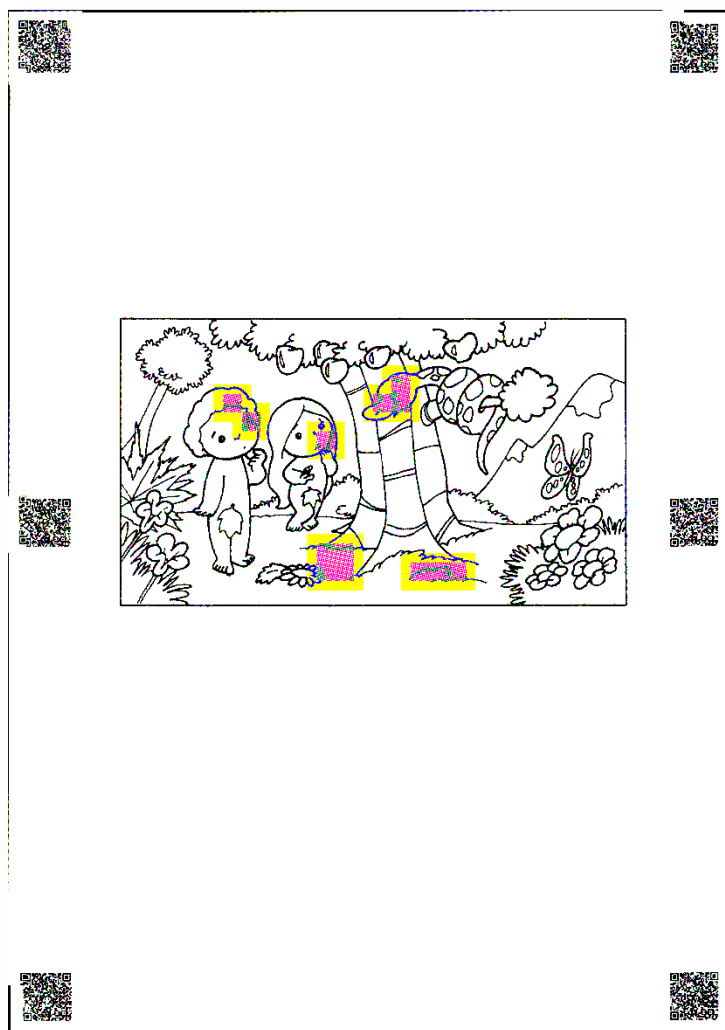


Figure 12: example of a verified graphic document, generated by the SIGNED prototype. The manipulated regions are highlighted.

3.10. System architecture

From the very beginning of the project, a high-level system architecture of SIGNED was defined taking into account all the requirements imposed to the solution and the different modules that would conform the final system, which are basically those described in the previous results. Thanks to the exhaustive specification work performed at the beginning of the project, the initial design proved to satisfy the needs for the final implementation of the SIGNED solution without the need to add significant modifications.

3.11. Security testing methodology

Result 11 consists in the specification of a testing methodology to verify whether the SIGNED solution (or future similar solutions) fulfils the requirements established at the beginning of the project. Such methodology defines the tests to be applied in order to check the functional and non-functional requirements of the whole SIGNED solution, defining for each test:

- The goal of the test
- Approach and operating methods
- Context (benchmark) for the test
- Plan of action (course) to execute/steps completed
- Expected result (outcome)

One of the most valuable aspects of the specified methodology is the part associated to the testing of the performance in detecting malicious manipulations, since it involves the construction and labelling of a database of test documents.

3.12. The SIGNED prototype

The first 10 S&T results described above were integrated to build the SIGNED prototype, which is a complete software demonstrator following the SaaS (Software as a Service) model.

The implemented prototype allows the use of the SIGNED solution in three different working modes, according to the requirements specified at the beginning of the project:

- The main solution is to hash the document, and embed the digitally signed hash in the barcodes of the securised document.
- The secondary solution is to use current **SecurePaper™** methods to embed the original document in the 2D barcodes, digitally signed, and use the hash algorithm, along with the channel estimation, equalization and comparison processes of the SIGNED solution to compare the hashes of the original and scanned documents. This moves all the heavy processing on the verification phase. Moreover, it might be the only valid solution in case the original document would be needed for processing by the end-user (as required by some national regulations). This solution doesn't break the main objective of the project: using a robust hash algorithm and a smart comparison algorithm to compare documents.
- The third solution, taken from the current **FiviDoc** implementation, is to calculate the hash of the document on the securisation server, and if it cannot be fitted into the number of barcodes agreed with the customer (which can be as little as two or three), upload it to a protected online repository. In this case both the hash of the document and the retrieval information must be digitally signed to assure integrity and authenticity. The verification process would verify the digital signature of the retrieval information and only if proven to be authentic would go to retrieve the hash of the document, and verify the digital signature. This solution breaks the principle of self-contained document, but it is necessary to satisfy

the demands of a significant group of end-users and meet the requirements of certain national regulations.

3.13. Main performance indicators

The tests performed to the SIGNED prototype have been carried out on the most widespread printer and scanner brands available on the market like Hewlett-Packard, Epson, Brother, Canon, Fujitsu and Xerox. The documents are subject to the distortions introduced by this hardware as expected, but the robust hash and the different mechanisms of the SIGNED solution for mitigating distortions do an excellent job at compensating them. The most important performance measures for the SIGNED solution are summarized below.

1. **Probability of Missed Detection (PMD) and False Alarm (PFA):** they are defined, respectively, as the probability of not detecting a fraudulent manipulation, and the probability of detecting an inexistent manipulation. Target values established for PMD and PFA are the following:
 - a. PFA below 0.001.
 - b. PMD: for a target PFA under 0.001, the following incremental (in order of difficulty) PMD values are targeted:
 - PMD1 < 0.001 for replacements of digits in arial 12 (or another typical font)
 - PMD2 < 0.001 for replacements of digits in arial 10 (or another typical font)
 - PMD3 < 0.001 for replacements of digits in arial 8 (or another typical font)
 - PMD4 < 0.001 for replacements of dots by commas in arial 12.
 - PMD5 < 0.001 for replacements of dots by commas in arial 10.

All the above target values are achieved by the final SIGNED prototype. The only remark must be done to PMD5, which is not fulfilled for arial font but for slightly bigger fonts like Verdana.

2. **Collision probability**, defined as the probability that different document regions (that look visually different) yield the same hash. This is crucial for the document authentication scenario envisioned in the SIGNED project, especially in high-security scenarios (e.g. financial, medical). The target value was established in 0.001, which is achieved by the final prototype.
3. **Minimum area size** where manipulations must be reliably detected: it measures the minimum area of a paper sheet where a manipulation can be spotted by the authentication system. Its target value was initially established in a square of 42x42 pixels on a 600dpi document. The SIGNED solution works with 64x64 pixels.

4. **Throughput:** it is measured as the number of pages (A4 size) that can be processed per unit of time, both by the securisation and verification modules. Thus, it is an indicator of the efficiency of these modules. The target values for this indicator are established in terms of the different PMD target values defined above:
 - 2 seconds per A4 page, for PMD1 and PMD2
 - 3 seconds per A4 page for PMD3
 - 4 seconds per A4 page for PMD4
 - 5 seconds per A4 page for PMD5

The final SIGNED solution fulfils the requirements for the securisation times. The verification times are larger due to the high computational complexity of some modules of the verification process, although there is room for further optimization, so as to achieve performance close to the target values.

5. **Size (length) of the document hash:** it is measured as the amount of information (in bits) needed to represent the digest of the document to be securised. This length was required to be as small as possible in order to be able to embed all this information in the document itself, although it was expected that for demanding PMD constraints it would be necessary to use fairly large hashes. Initially, the target value for the hash size was 4 Kilobytes. For typical documents, the final SIGNED solution provides hashes between 4.8 and 12 KBs for a medium security levels, whereas for very high security levels (detection of very small modifications) it requires from 39 to 170 KBs. These hash sizes mean that fewer documents can carry the hash within the barcodes, but there are several solutions around this, keeping the robust hash is still one of the building blocks of the securisation and verification process. During the project, for example, the Italian regulations stated that the entire document must either be carried along the printout in the barcodes of the securised document, or must be available for retrieval by the end user. That change would forbid to use the robust hash itself in the barcodes, thus giving rise to the need of the secondary working modes envisaged in the project.
6. **Compatibility level with existing printers and scanners.** The SIGNED solution was tested with documents generated with the most widespread printer and scanner brands available on the market (Hewlett-Packard, Epson, Brother, Canon, Fujitsu and Xerox). The detection capabilities of SIGNED were proved robust to middle-quality and high-quality scanners.

3.14. Conclusions

It can be concluded that the final SIGNED solution successfully addresses the main technical limitations identified at the beginning of the project in other existing solutions for document authentication:

1. Lack of robustness. Existing robust hashing schemes have been mainly designed to support slight compression and very minor geometrical changes. The techniques designed in the SIGNED project robust to digital-analog (print) and analog-digital (scan) conversions, and they proved to be valid for a wide variety of scanner and printer models, hence extending the capabilities of the preceding solutions Fividoc and SecurePaper, which were considered the most competitive solutions to date.
2. Lack of universality. Most existing document authentication methods are focused in specific document formats, hence limiting their scope and possibility of application to new scenarios. The SIGNED authentication system is applicable to any kind of printed document, regardless of its format and layout. Hence, it is applicable to documents containing only text, both text and images, document IDs, certificates, etc., extending the capabilities of current Fividoc and SecurePaper solutions.
3. Need of human intervention and inefficiency. Some existing document authentication methods require human intervention at some point of the document generation or verification process. Some solutions, for instance, simply convey in the barcodes one link to the original document, which must be compared by visual inspection to the document under verification. This is unacceptable due to the high time consumption and the high probability of making mistakes, besides posing privacy issues in the case of confidential documents. The SIGNED verification process is completely automated, being able to securise documents and checking their authenticity in a few seconds.
4. Lack of security. Existing document authentication solutions do not provide in paper documents similar security levels as the established Digital Signature (DSAS) technique, that the SIGNED project is indeed able to achieve.
5. Most of the previous and current existing solutions for document authentication have a significant negative impact on cost, as they require specific types of paper, printers and/or scanning/analysis devices. This prevents their wide deployment and often requires trained staff to do the securisation and verification tasks.
6. Sharing of private keys. Classical watermarking schemes for document authentication require the sharing of a private key between the sender and the receiver, thus implying the need of a secure scheme for transmitting it. The SIGNED authentication system relies on Public Key Infrastructure to overcome this problem.

4. Potential impact

In spite of the important advance through an electronic world, we still live and will be living during the next years in a paper-based world. Documents as identity cards, passports, driving licenses, academic degrees and certificates, medical records, titles of ownership, licenses, labels, contracts, checks, birth, marriage and death certificates, invoices, etc, still will be used in printed form for many years. Any of those documents, either forged or fraudulently altered, allow criminals to obtain public benefits, fraudulently practice a profession, commit identity fraud, cheat, steal and sell goods, counterfeit products, get funds from a bank, etc.

Nowadays, the most widely extended method for checking the validity of printed documents is the mere visual inspection of the document to be validated and a trusted copy. This is a highly inefficient process, since it consumes a lot of time and human resources. Moreover, this manual validation is not always possible because the original trusted document is often not available, and it is largely prone to human errors, especially when a big volume of documents must be validated in a short time.

Usually, the authenticity of a printed document is taken for granted upon the existence of a hand-written signature or some sort of stamp, and people do believe in the authenticity and the integrity of these documents even if it is the first time they see that handwritten signature. In documents like IDs or academic degrees, physical-chemical-optical security measures are applied such as fluorescent inks, watermarks and holograms. In theory, these measures provide a high level of security to the protected documents, but they present some important disadvantages with regard to the SIGNED technology:

- The validation of the documents must be performed with special (and costly) hardware devices, and in general the application of certain experts' personal criteria to do the final decision about authenticity and integrity.
- The validation is not automatic.
- Most of the times the securisation process cannot be performed without using specific, costly devices.
- The protection of common documents such as certificates, payslips or bank statements, can only be made with low security measures and even so the cost of this protection is very high.
- These measures cannot be used for protecting the documents issued by e-Government services, which are downloaded by the citizens to be printed with their own common printers over their own paper.

So far, security document technologies were aimed at creating materials that were difficult to be imitated with enough quality to go undetected by experts using specific instruments. This is an evidence-based approach, and therefore it is not designed to ensure an effective and efficient detection of counterfeits. It should be realized that most documents are usually handled by unskilled people, which rarely belong to the issuer entity. Therefore, most of receivers of the documents are unaware of the usage of such security measures on the documents, and usually they are poorly motivated to detect the faked ones. Usually, the aim of the counterfeiters is not to obtain the "perfect forgery", as what they really need is to achieve

fake documents with enough quality to go undetected by the receivers. It is necessary to use new security technologies that lead to efficient and effective validating processes, even if unskilled receivers perform these validations. Furthermore, these new technologies must achieve high detection rates independently of the user motivation, and must avoid the loss of productivity due to the visual comparison of documents. This is exactly what SIGNED technology provides.

The lack of effective measures causes billionaire direct and indirect losses worldwide, as nearly all committed crimes are preceded and covered up by fake documents. In addition, the problem has been increasing in recent years, as far as it is connected with the availability of sophisticated yet easy to use tools for manipulating digital images: nowadays, in fact, a paper based document could be easily manipulated by digitalising it (obtaining an image through scan), manipulating the image through image editors, and then printing it; the printed document will look authentic to users, and no easy-to-use systems are available to verify the authenticity of such document. To better describe the problem, a very simple scenario related to possible frauds with paper-based documents is reported:

Fraud scenario: how to reduce income taxes by manipulating bank paper documents

In most European countries citizens are entitled to deduct, from taxable incomes, bank mortgage interests related to home acquisition. A very simple fraud could be realised:

1. The bank sends the document to the citizen attesting a mortgage interest (for example of 3.000 Euro), to be used for income tax declaration.
2. The citizen scans the document, and manipulates it increasing the mortgage interest to 8.000 Euro
3. Then he uses the manipulated document for its own tax return, avoiding paying taxes on 5.000 Euro of income.

Similar frauds could be easily realised by companies that can manipulate leasing or bank interest rates to reduce their EBIT (Earning Before Income Taxes). Many other similar frauds could be also realised manipulating medical documents (for insurance companies or for tax deduction), etc.

The kind of frauds described in the presented scenario are very easy to be realised and, unfortunately, very difficult to be detected: as an outstanding example, during the Parmalat crack of recent years, it was discovered that frauds were conducted just by scanning the logo of a famous US bank and then using such forged documents to prove non-existing company credits.

Though most of the document frauds are usually never detected and therefore not quantified, we can have an idea of the impact of this kind of fraud by analysing some reports of markets

with strong investments in fraud prevention. For example, in the 2011 ABA Deposit Account Fraud Survey we can find that 73% of banks reported check fraud losses in 2010, which amount for approximately \$893 million.

There is the need, therefore, of a system able to authenticate paper based documents; this means, to have a system that is able to recognise if a certain paper based document has been manipulated after being issued and/or after being verified (sealed). Furthermore, the new system must:

- be able to be used by unskilled and unmotivated people.
- be able to work with common printers and scanners.
- be able to validate a high amount of documents in an automated manner.
- be integrable into the systems and processes of the entities.

The project final outcome is a new authentication system for paper based documents, based on robust image hashing technologies and new comparison algorithms. SIGNED is composed by a securing module and a validation module, than can be accessed by its SOAP web service interfaces, and a web application that allow to use SIGNED without the need of performing integration with other systems.

The process of using SIGNED is very simple, and summarised below.

- The documents must be secured before being issued. The tasks performed during this securisation process are dependent on the working mode, but SIGNED always insert one or more QR barcodes into the securised document.
- The securised electronic document is issued either in printed form or in electronic form. If it is issued in electronic form, its owner can print it using his/her own printer.
- The secured document is delivered to a third entity, which needs to validate its authenticity and integrity. This entity must scan the document and use the validation module of SIGNED. With respect to the integrity, SIGNED will apply its algorithms to detect alterations on the content not due to the typical distortions of the Print&Scan channel. The fraudulent detected alterations will be highlighted over the scanned document, so a visual evaluation of its importance can be performed.

The main features of the SIGNED solution are the following:

1. **Simplicity of integration.** The SIGNED solution was specifically designed to be integrable in the back-end systems of both the issuing entities and the validation entities. Third applications only have to invoke the SIGNED SOAP (Service Oriented Architecture Protocol) web services interfaces to use the securing or the validation functionalities.
2. **Simplicity of usage.** As the SIGNED solution can be integrated into the current systems of the organizations, its use can be transparent to the user. The validation of a document with SIGNED is as simple as scanning a document and sending the scanned file to the validation module of SIGNED, which will perform the validation automatically informing about the alterations found on its content.

3. **Compatibility with common printers and scanners.** The SIGNED algorithms were specifically designed to work with common printers and scanners. Tests were performed with devices of most of the biggest manufacturers.
4. **Applicability to a large amount of documents.** As SIGNED is a software tool, both the securing and the validation processes can work in batch mode.
5. **High capability for detecting fraudulent alterations.** The performed tests show that the Detection Rate of minimum alterations as the replacement of a digit in Arial font with size as little as 8 points is higher than 99.9%, keeping the False Positive Rate below 0.1%.
6. **Universality**, i.e. applicability to all type of documents. SIGNED algorithms are image-hashing based. No OCRs (Optical Character Recognition) are used. Thus, SIGNED can be applied to documents with text but also with images as logos, stamps or handwritten signatures.
7. **Privacy-preserving.** Currently, document validation in several scenarios is done by human intervention, visually comparing the document to be validated and the original digital document, which may contain personal data. However, in most EU Member States, the access (understood as viewing or downloading) citizen's personal data is restricted by privacy regulations. SIGNED does not need to show the original document, and therefore is respectful with the privacy legislation in force.
8. **Accurate detection of the alteration location.** SIGNED spots the locations of all the manipulations found in the document.
9. **Detection of photocopying.** SIGNED includes a module to estimate if a scanned document has been printed only once or more. This functionality provides another tool to prevent document fraud, allowing to distinguish copies from original documents.
10. **Authenticity based on PKI standard.** The digital signature of the information embedded in the securised documents, using standard x.509v3 certificates and OCSP servers, guarantees the authenticity of the documents.
11. **Different security levels.** Three levels of security can be used, each one corresponding to a different working resolution: 150 dpi, 300 dpi and 600 dpi.
12. **Different working modes.** Three different working modes can be used, which can be set in the securisation stage. In the first working mode, the digitally signed hash of the document is stored into the barcodes of the securised document. In the second working mode, the original digital document is signed and stored into the barcodes. In this case the hash of the original document is computed in the verification stage. In the third

working mode, the hash is computed in the securisation stage and stored in a database, and only a unique document identifier is stored in the barcodes.

The project results will enable SMEs, Public Administration, banks, insurance companies and other stakeholders, facing every day the problem of paper based document authentication, to deal with important documents like financial documentation and legal documents in a trustworthy and reliable manner. In fact, the technology developed within SIGNED will be applicable to any exchange of sensitive information on paper-based documents. The possibility of having an authenticity certification seal within the paper based document will guarantee that SIGNED will not be specific to a given industrial domain or field; it will rather be a major achievement for the benefit of all industries, service companies, public administration and citizens.

The applications resulting from the SIGNED technology are numerous. This technology may be used for a wide variety of purposes, some of which being:

- Industrial Applications: exchange of information between customers/suppliers/partners and between industries and citizens could benefit from the project's success. Information such as corporate balance sheet, bank documents, certifications, legal authorizations, etc., could be exchanged through different companies in a certified way, meaning that each company will have the possibility to check the authenticity of the received documents. Organisms like the REA or the European Commission can benefit from SIGNED for instance for performing an automatic validation of the forms C received from the beneficiaries of funded projects.
- Application complementary to Digital Signature: whereas the classical Digital Signature Standard (DSAS) is only applicable to digital files, SIGNED will extend this concept to paper documents, thus filling the gap, i.e. a document protected with SIGNED will be protected either in digital or in printed form. For example, in the context of future e-administration services, the citizens will be able to request, download and print at home any official document that could be used later on as an original and authentic document in legal transactions or in further transactions with the Public Administration, e.g. for tax declaration. Hence, the SIGNED technology will favour the modernization of the Public Administrations in terms of e-government without impairing other organizations that cannot adopt this kind of advanced solutions. Many more people could enjoy the DSAS's benefits and a larger scale deployment of DSAS become feasible.
- Applications for the financial sector: SIGNED technology will significantly contribute to reduce the large economic losses in the financial sector due to check fraud, internal fraud, or fraud of loans granted to people who present forged documents like IDs, payslips or social security documents.
- Identity fraud: most of personal identity documents such as birth certificates, wedding certificate, etc, are still paper based. While the most developed countries are moving towards digital documentation for important documents such as passports, most people still use and will continue using for the forthcoming years paper based documents, that

need to be checked at country boundaries or that are used inside the European Union to provide new identification documents. The possibility to check the authenticity of a paper based document will reduce frauds and will simplify the work of people responsible for checking the authenticity of certain documents and identifying terrorists or criminals travelling from one country to another.

In addition to reduce the losses due to document fraud, SIGNED will allow to the organizations to make some of their procedures much more productive. Just as an example, we can describe the process followed by the REA during the financial reporting of the projects. In each reporting period of a project, each beneficiary must to fill in the Participant Portal a Form C indicating its costs during the reporting period. This Form C must be validated by the coordinator, then printed, sealed, signed by hand and sent by postal mail to the Commission. The Commission must verify by visual comparison that the received original printed version has exactly the same content as the electronic version stored in the Participant Portal. The target of this comparison is not to avoid fraud, but mistakes. With SIGNED, the Form C could be secured before being printed, and the REA could introduce all the received Forms C into a scanner, and validate automatically that each of the printed Forms C are identical to its electronic version. Similar processes can be performed in other industries, such as insurance companies or online banks, which in many cases send the contracts to the customers who must sign them and send them back to the companies, where some people have to perform visual comparisons with the original version of the contract in order to verify that the content of the contracts has not been altered.

4.1. Dissemination and exploitation activities

Several types of dissemination activities were performed as publication of articles in magazines and newspapers, participation in forums, congress and seminars, or mentions in the beneficiary's websites and through the social networks. However, the most relevant dissemination activities were:

- The large number of potential final clients and resellers informed about SIGNED by the SMEs participating in the project, not only in European countries but also in Latin American countries as Brazil, Mexico, Brazil, Colombia, Peru and Ecuador.
- The showcasing of the SIGNED prototype at the industrial demonstration session of the 4th IEEE International Workshop on Information Forensics and Security (WIFS 2012) held in Tenerife, Spain, December 2-5, 2012. WIFS is the flagship conference of the IEEE on information security, and major actors of the multimedia security industry and the scientific community were present. Link: <http://wifs12.org/>. During the demonstration people could test the SIGNED prototype, performing validations of authentic documents and fake ones prepared by the Consortium beforehand, and also could perform fraudulent alterations of the content of the authentic documents before validating them. People got surprised with the great detection capability of SIGNED.

5. Annex I – Acronyms and abbreviations

DSAS Digital Signature

PFA Probability of false alarm

PMD Probability of missed detection

QR Quick Response 2D barcode

SaaS Software as a Service

SOAP Service Oriented Architecture Protocol


6. Annex II – Contact sheet





www.signedfp7.eu

Coordinator and contact: Alberto Malvido García (alberto.malvido@bitoceans.es)

Participant SMEs:

	Bit Oceans Research S.L. (www.bitoceans.es)
	LAND S.R.L. (www.land.it)
	Global Security Intelligence (www.globalseci.com)

Participant RTDs:

	GRADIANT - Galician R&D Centre in Advanced Telecommunications (www.gradiant.org)
	Università degli studi di Roma “Tor Vergata” (www.uniroma2.it)